# Regulation of Eukaryotic DNA Replication:
## Genome Architecture & Polymerase Switching

Phoebe Oldach

University College, Oxford
Sir William Dunn School of Pathology, Medical Sciences Division

A thesis submitted to
the University of Oxford for the degree of
DPhil in Chromosome and Developmental Biology




Hilary Term 2021
Word Count: 45,097

# Abstract

Eukaryotic DNA replication is tightly regulated, requiring the coordinated action of many different cellular processes and proteins to facilitate duplication of the entire genomic template.

This work investigates two different aspects of this coordination: the timing regulation that leads some genomic loci to replicate early in S phase and some late, and the enzymatic regulation that allows nascent strand synthesis to be carried out in an efficient and error-minimizing way by one of a family of distinct DNA polymerases.

First, the role of 3D genome architecture in human cell replication timing is analyzed with the use of a system for inducible cohesin degradation. Loss of cohesin has been shown to drive loss of topologically associating domain (TAD) structure across the genome. Thus, high throughput sequencing and microscopy were used to assess replication timing differences between wildtype and cohesin-ablated S phase conditions. Neither acute S phase cohesin ablation nor sustained ablation from G1 through S phase resulted in genome-wide changes to replication timing domains (RD). Furthermore, specific genomic loci hypothesized to show a replication timing response to cohesin loss – such as TAD boundaries, CTCF binding sites, and genes that showed differential expression in the absence of cohesin – showed no significant changes. Thus, it was concluded that cohesin is necessary for neither the establishment nor execution of replication timing domains.

Second, the potential for switching during ongoing leading strand synthesis by polymerase epsilon is investigated in budding yeast. A system was developed to incorporate (1) an inducible and tagged polymerase epsilon Pol2 subunit in addition to the endogenous Pol2 subunit, (2) a long replicon free of licensed replication origins, (3) robust cell cycle synchronization, and (4) a quantitative ChIP-Seq method for detection of the induced and endogenous Pol2 subunits on chromatin. Preliminary experimentation utilizing this system suggest an absence of rapid and widespread polymerase epsilon switching during ongoing replication. This opens the door for future experiments assessing polymerase switching in response to replication stress and obstacles such as UV-induced DNA damage.

## Coronavirus Impact Summary

The emergence of covid-19 has had immeasurable impacts on life and livelihood the world over. On the smallest scale of its disruptions falls my DPhil research. While I am very grateful to have thus far remained healthy and productive over the course of this pandemic, the implementation of a nationwide lockdown and the closure of University facilities to non-Covid research in March 2020 has changed the trajectory of my DPhil work as described below.

Brief Summary of Work Planned for Completion before Lockdown

My second results chapter, 'Polymerase Exchange,' sets out the creation of a system to assess polymerase epsilon switching *in vivo* in *S. cerevisiae* through high throughput sequencing.

Pilot ChIP-Seq experiments employing this system suggested that further control over cell cycle progression and replication initiation would be beneficial. Thus, work was initiated to engineer a temperature sensitive *cdc7-1* mutation into the experimental strain. This strain creation was interrupted by the Covid lockdown.

The polymerase induction system was validated in bulk populations through Western blotting. Pilot ChIP-Seq experiments showed evidence of potential leakiness in the Pol2 induction; this suggested that assaying protein induction in single cells through a fluorescent reporter gene and flow cytometry would be valuable for characterizing the heterogeneity of the induction system. This strain creation was interrupted by the Covid lockdown.

Beyond these specific aims actively interrupted by the Covid lockdown, the project aims following establishment and validation of the induction system included the interrogation of polymerase switching under perturbations hypothesized to upregulate switches (UV-induced DNA damage and increased concentrations of the induced polymerase). Due to various disruptions and the vagaries of experimentation, these aims have not been achieved in the course of this DPhil work. However, they – along with other future directions laid out in the Discussion portion of the Polymerase Exchange chapter – are now within reach for future researchers given the development of the experimental system and pilot experiments described herein.

# Acknowledgements

I am very grateful for the support and camaraderie of the Replication Group. In particular, the scientific advice, administrative support, and cheerful resilience from Carolin Müller has been invaluable to my progress through this DPhil.

I would like to thank the Dunn School for its enriching scientific community and administrative support. I appreciate all of the time and openness that Head of Department Matthew Freeman was able to share; the frank conversations and guidance from H.R. Manager Andrew Souter; the action for concrete change in Medical Sciences Division policy from Director of Graduate Studies Anton van der Merwe; and the substantive logistical help from Lucinda Risius and the entire Dunn School finances team. Outside of administration, I deeply appreciate the way that so many in the Dunn School community contributed to helping the Replication Group navigate challenging situations. In particular, thanks are due to Liz Robertson for going to bat for us and Liz Bikhoff for her sixth sense for when and where our lab meetings were being held and her seemingly endless supply of cookies. My secondary supervisors Dr. Carolin Müller, Professor Jordan Raff, and Professor Chris Norbury were especially helpful in providing me with scientific and strategic advice throughout the latter half of my DPhil.

I am also very grateful to University College and the guidance of its Graduate team, in particular my college advisor, Professor Peter McHugh, and the Dean of Graduates, Professor Caroline Terquem.

My work, like all science, depended on the work of many other scientists. I am very grateful for the direct assistance of Line Eriksen and Michal Maj in the Dunn

School Flow Cytometry facility, and Amanda Williams and Rebecca Busby in the Zoology Sequencing Facility. I would also like to thank Masato Kanemaki for sharing the SCC1-AC Tir1 HCT116 cell line, and Anne Donaldson for sharing the origin-deletion *S. cerevisiae* cell lines. I am also grateful to the Carr lab, in particular Alice Budden and Karel Naiman, for sharing their expertise and friendliness when I traveled to the University of Sussex to learn Pu-Seq with them in November, 2018.

# Abbreviations

| | |
|---|---|
| ACS | ARS (autonomously replicating sequence) consensus sequence |
| AID tag | auxin inducible degron tag |
| ARS | autonomously replicating sequence |
| bp | base pairs |
| BrdU | bromodeoxyuridine / 5-bromo-2'-deoxyuridine |
| ChIP | chromatin immunoprecipitation |
| chr | chromosome |
| CI | confidence interval |
| CMG | Cdc45-MCM-GINS helicase |
| CPD | cyclobutene pyrimidine dimer |
| CTD | C terminal domain |
| ER | estrogen receptor |
| ERCE | early replicating control element |
| FACS | fluorescence activated cell sorting |
| FDR | false discovery rate |
| FRAP | fluorescence recovery after photo-bleaching |
| GCR | gross chromosomal rearrangement |
| GINS | Sld5, Psf1, Psf2, Psf3 |
| HBD | hormone binding domain |
| HSP | heat shock protein |
| HRP | horseradish peroxidase |
| HTS | high throughput sequencing |
| HU | hydroxyurea |
| IAA | indole acetic acid / auxin |
| IF | immunofluorescence |
| IP | immunoprecipitation |
| kb | kilobase |
| MCM | minichromosome maintenance protein complex |
| MLN | multilobed nucleus |
| MMS | methyl methanesulfonate |
| Mrc1 | mediator of the replication checkpoint 1 |
| NTD | N terminal domain |
| $OD_{600}$ | optical density at 600 nm |
| ORC | origin recognition complex |
| PCNA | proliferating cell nucelar antigen |
| PI | propidium iodide |
| Pre-LC | pre-loading complex |
| Pre-RC | pre-replication complex |
| Pol | polymerase |
| Pu-Seq | polymerase usage sequencing |
| RFC | replication factor C |
| RT | replication timing |
| RD | replication timing domain |
| siRNA | small interfering RNA |
| SPT | single particle tracking |
| TAD | topologically associating domain |

| | |
|---|---|
| TDP | timing decision point |
| ts | temperature sensitive |
| WCE | whole cell extract |
| wt | wildtype |

# Table of contents


CHAPTER I. INTRODUCTION 12
CHAPTER II. MATERIALS AND METHODS 20
Additional resources 21
Molecular biology 21
Bacterial culture and manipulation 34
Ampicillin selection for *E. coli* 35
Chloramphenicol selection for *E. coli* 35
Yeast culture and manipulation 36
Yeast-Peptone-Adenine-Dextrose (YPAD) medium 37
Yeast-Peptone-Adenine-Dextrose (YPAD) agar 37
YPA 37
Sporulation plates 37
G418 selection medium 37
CloNat selection medium 38
Raffinose 38
Galactose 38
Estradiol 38
Human cell culture and manipulation 44
BrdU replication foci 47
SCC1 47
Computational analysis 49
Fork age 55
Polymerase localization 55
CHAPTER III. COHESIN-MEDIATED GENOME ARCHITECTURE DOES NOT DEFINE DNA REPLICATION TIMING DOMAINS 57
Introduction 58
Genome replication and architecture 58
Genome architecture as a determinant of replication timing 59
Implications of replication timing 62
Cohesin 64
Auxin degron tag 66
Motivation 69
Results 70
System development and validation 70
Efficient and rapid degradation of endogenous SCC1 70
S phase progression in the absence of cohesin 71
Establishing the Repli-Seq protocol 73
Loss of SCC1 does not perturb replication timing patterns 75
Asynchronous depletion of SCC1 75
Depletion of SCC1 from early G1 82
Impact of genome organization on replication timing 89
Discussion 91
Perspective and Further Work 92
CHAPTER IV. INVESTIGATION OF POLYMERASE EPSILON EXCHANGE AT THE *S. CEREVISIAE* REPLICATION FORK 97
Introduction 98
Eukaryotic genome replication 98
Polymerase epsilon structure and regulation 100

Polymerase epsilon expression and dynamics 104
Polymerase epsilon roles in S phase 106
CMG (Cdc45 – MCM – GINS) helicase maturation 106
CMG progression and replisome speed 108
Replisome fidelity 111
Replication checkpoint involvement 112
$RFC^{Ctf18}$ 115
Mrc1 116
Replication-coupled nucleosome assembly 117
Response to lesions 119
Dynamic processivity 121
Polymerase epsilon exchange 123
Aims 126
**Results** **128**
Motivation and scope 128
System development 128
Inducible Pol2 129
Localization control 129
Expression control 135
Gal2 knockout 137
Long replicon 139
Fork age modeling 140
Long replicon scenarios 142
Chromosome 6 right 142
Time restricted initiation 143
Temperature-sensitive *cdc7*, galactose-inducible POL2-Flag strain creation 144
Synchronization 146
System characterization 155
Speed of induction 155
Quantification of induction level 158
Cell cycle impact of Pol2 overexpression 160
Titrating Pol2 expression 162
Population heterogeneity in galactose induction of Pol2 165
Cell cycle switching experiments 166
Arrest in HU following constitutive induction or repression 166
Arrest in HU with non-induction or induction 170
S phase progression with non-induction or induction 174
Slowed S phase progression (23°C) with non-induction or induction 177
**Discussion** **182**
Creation of a regulatable polymerase system in *S. cerevisiae* 183
Cell impacts of polymerase epsilon overexpression 184
Assaying for switching behavior 186
Future directions 188
Alternative read-out 188
Switching to other polymerases 191
CHAPTER V. DISCUSSION 192
APPENDIX 195
REFERENCES 220

# Chapter I. Introduction

Implicit in the term ‘semiconservative’ is the partially destructive nature of DNA replication. To duplicate the genome, duplex DNA must be unpackaged and unwound so that each naked strand can act as a template for the synthesis of a reverse-complimentary partner. While critical for the perpetuation of cellular life, this process through which one duplex becomes two can pose significant risks to genome integrity [1–4]. From topological stress driving double strand breaks and then rearrangements, to replicative slippage or switching introducing copy number variations, to unrepaired synthesis errors cementing into mutations: the process of genome duplication can be a powerful driver of genome alteration.

Unfaithful replication is fundamental to evolution, as these variations form the basis for selection. However, a careful balance must be struck; sufficient replication fidelity to propagate successful genomes is necessary for selection to be meaningful. And while the overall consequences of unfaithful replication can be advantageous to a species, they are frequently deleterious for an organism. They can result, for example, in the attenuation or total loss of gene function, or conversely a gene’s hyperactivity. In either direction, these perturbations can lead to the unfettering of cellular signaling and processes, driving a cell either quickly towards death or towards uncontrolled and unsustainable proliferation, as in cancer. As the consequences of unfaithful replication can be profound for a cell or for an organism, so too is the scale of the regulations evolved to safeguard the process.

These regulations cover many aspects of replication, from checkpoints modulating entry into and progression through S phase to multi-protein mechanisms working alongside the replisome. They orchestrate everything from the removal or

facilitated-bypass of replication obstacles to the rapid repair of synthesis errors and the chaperoning of epigenetically-marked histones back to the appropriate genomic loci. They are instrumental in safeguarding replication fidelity, while also driving replication speed. Of the many facets of regulation integral to preserving the genome and ensuring efficiency throughout DNA replication, two are investigated in the scope of this DPhil work: the control of replication timing in human cells and the stability and switching of the synthesizing polymerase in the replisome of budding yeast. An overview of eukaryotic replication process and machinery is provided here, followed by more in-depth introduction to the specific workings of replication timing in human cells and polymerase switching in budding yeast in their respective chapters (Chapters III and IV).

## Eukaryotic DNA Replication

### *Licensing*

Preparations for replication begin in G1 phase as the origin recognition complex (ORC) forms a pre-initiation complex (pre-IC) with licensing factor Cdc6 on DNA loci termed 'origins of replication' (origins) [Figure 1] [5]. Two mini-chromosome-maintenance complex hexamers (Mcm2-7) are then recruited to each origin with the help of Cdt1 [5]. The Mcm2-7 complex makes up the catalytic core of the replicative helicase, powered by two terminal AAA+ ATPase domains on each subunit. The complex remains inactive throughout the licensing stage in G1 as its activity is dependent upon activation by two S phase kinases, Dbf4-dependent kinase (DDK) and CDK [5].

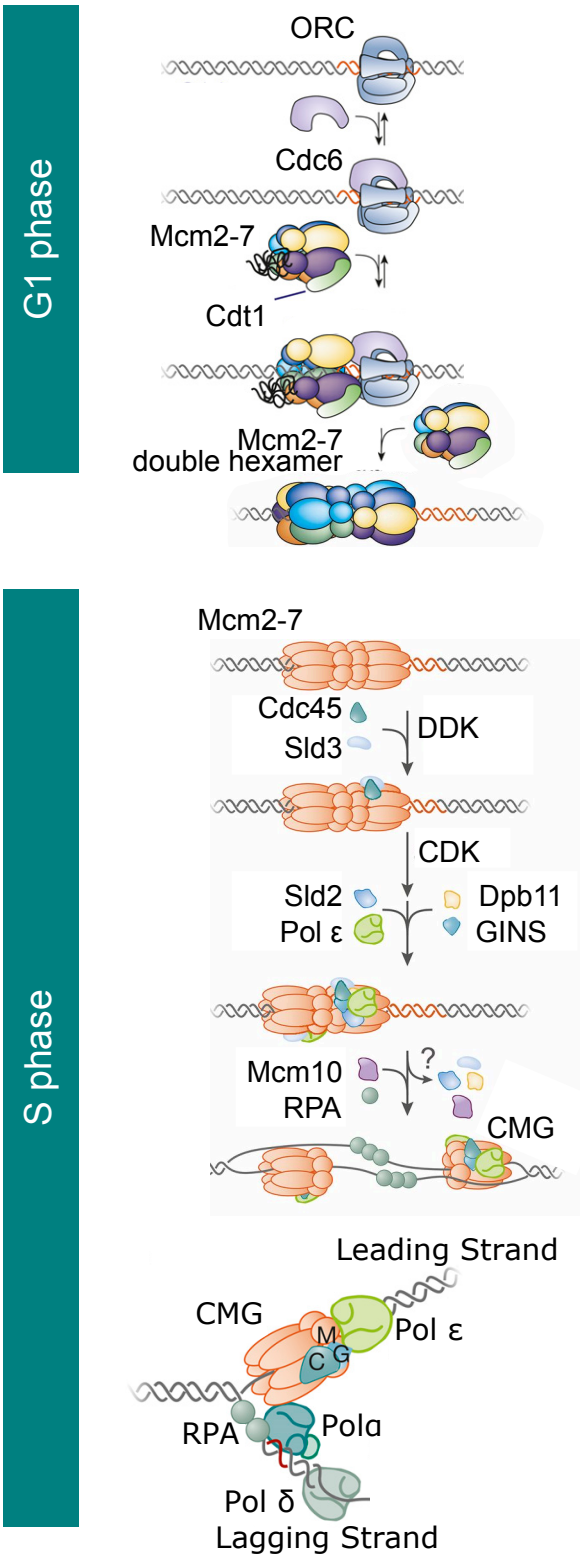



**Figure 1. Replication initiation and elongation schematic.** Replication licensing occurs in G1 phase and is driven by ORC binding to origin DNA, followed by the recruitment of two hexamers of Mcm2-7 through loading partners Cdt1 and Cdc6. Replication initiation occurs in S phase, as it depends upon two S phase kinases, DDK and CDK. DDK phosphorylations drive the recruitment of Cdc45 and Sld3, and subsequent CDK phosphorylations drive recruitment of the pre-loading complex, comprised of Sld2, Dpb11, GINS, and polymerase epsilon. Association of Mcm10 allows

for final CMG activation, and RPA coats the resulting single stranded DNA. Replication elongation proceeds with the two CMG helicases loaded at each origin progressing in opposite directions. DNA synthesis is primed on both strands by polymerase alpha, and polymerase epsilon carries out the bulk of the leading strand synthesis while polymerase delta carries out the bulk of lagging strand synthesis. Figure adapted from [5].

In budding yeast, this pre-IC assembly happens at precisely-defined genomic loci, as the budding yeast ORC binds specifically to an AT-rich motif termed the autonomously replicating sequence (ARS) consensus sequence (ACS) [5]. By mutating the ACS sequence, it is possible to specifically ablate the function of an origin of replication in the *S. cerevisiae* genome [5].

In other eukaryotes, origin determination is not as clear – no conserved ORC-binding motif has been identified [6]. Rather than a specific sequence, ORC-binding has been shown to be driven by DNA characteristics: origins have been associated with active promoters, open chromatin, CpG islands, and secondary structure elements such as G-rich repeat elements that are able to form G quadruplexes [6]. The three-dimensional organization of the genome has also been associated with origin selection: origins were found to be enriched for binding of the cohesin protein, which defines chromatin loop domains [7].

### *Initiation*

Initiation begins with phosphorylation of the Mcm2-7 complex by DDK. This drives a conformational change that leads to the recruitment of replisome proteins Cdc45, Sld3 (Treslin in humans), and Sld7 (MTBP in humans) [8]. CDK then acts additionally at the pre-RC to phosphorylate Sld3. Off of the chromatin, CDK phosphorylation helps to drive the formation of the pre-loading complex (pre-LC), which is comprised of CDK-phosphorylated Sld2 (REQL4 in humans), polymerase epsilon, Dpb11, and the GINS complex [9].

After assembly, the pre-LC loads onto the pre-RC, bringing GINS into contact with the MCM and Cdc45 to form the CMG (Cdc45-MCM-GINS) helicase [10]. GINS is required for replication elongation and has a primary function in holding the MCM and Cdc45 together [10]. Mcm10 is further recruited as a final step in CMG activation, at which point the DNA at the origin is melted and the CMG switches from encircling double stranded DNA to wrapping around a single strand [5]. The two CMGs at each origin can then begin to progress in opposite directions, with each proceeding 3' to 5' along the encircled strand [11]. CMG progression melts the duplex DNA, exposing a single stranded template for first RNA-DNA priming by the polymerase alpha primase complex, and then synthesis by polymerase epsilon and delta [8].

### *Initiation timing*

In budding yeast, initiation of each origin happens at a precise and reproducible time in S phase, which is dictated by the time at which each origin gained access to the initiating factors such as DDK, present in the nucleus at limiting levels[12,13]. In human cells, while origins are not so precisely mapped, limiting-factor dependent activation also occurs at different characteristic times throughout S phase, resulting in reproducible patterns of replication timing across the genome [14].

In addition to initiation occurring asynchronously across the genome, there are some origins at which replication initiation does not occur in a given S phase. Termed 'dormant origins', these can be activated in case of stressed replication or specific obstacles leading to fork stall or collapse[15]. Thus, dormant origins act as safeguards to ensure the efficient completion of DNA replication.

### *Elongation*

The two single strands of the unwound DNA duplex provide two distinct templates for replication. The ‘leading strand’ is the strand unwound in a 3’ to 5’ direction, which provides a straightforward template for continuous 5’ to 3’ synthesis by a DNA polymerase following the direction of the CMG progression. The partner strand, or ‘lagging strand’, is unwound in a 5’ to 3’ direction, and thus its duplication requires discontinuous 5’ to 3’ synthesis in the direction opposing CMG unwinding. Products of this discontinuous synthesis are termed Okazaki fragments, and they are ligated together post synthesis [16]. In eukaryotes, three B-family DNA polymerases are responsible for the bulk of genome duplication, polymerase alpha, epsilon, and delta [17]. In both budding yeast and human cells, polymerase epsilon has been shown to carry out the majority of leading strand synthesis, while polymerase delta has been shown to carry out the majority of lagging strand synthesis, along with playing a role in synthesis of both strands at initiation [18].

### *Termination*

The convergence of two opposing replisomes marks the termination of their replication. At this point, the forks are dissembled from the DNA through a process that involves ubiquitylation of Mcm7 to safeguard genome integrity and prevent re-replication [5].

## Thesis Overview

While the steps outlined above give an overview of the essential components and key steps required for DNA replication in budding yeast and human cells, there are manifold additional factors and regulatory systems in place that overlay on this

process. The control of replication timing and polymerase switching act in such a way, modulating the process of replication to ensure its efficiency, fidelity, and robust adaptation to a dynamic cell environment.

Work to further unpick the mechanisms of these – and other – replication modulators will help not only to understand the fine-grained orchestration of replication and the interplay between regulatory systems, but also to elucidate the overall limits of the replication program. By understanding the variations that replication regulatory systems can respond to, and the constraints of these systems' adaptations, it will be possible to develop a deeper understanding of how the process of genome duplication goes awry, leading to genomic variation and disease. The work of this thesis is presented in two chapters, which each have Introduction, Results, and Discussion sections specific to the experimental questions of that chapter. There is one consolidated Materials and Methods chapter covering the techniques used in both chapters.

# Chapter II. Materials and Methods

## Additional resources

For ease of reference, some details are compiled in tables available in the appendix.

The composition for buffers written in bold throughout these Methods can be found in the Buffer and Reagent Table.

Genotypes and engineering details for yeast strains used or created in this work can be found in the Yeast Strain Table.

Plasmids resources created in this work are detailed in the Plasmid Table, and the sequence of primers used are listed in the Primer Table.

## Molecular biology

### 1. Polymerase chain reaction (PCR)

PCR reactions were carried out with either Phusion High Fidelity (NEB, Cat #M0530S) or BioTaq (Meridian, Cat #BIO-21040) polymerase. In general, the Phusion proofreading polymerase was used for reactions in which synthesis fidelity was important (*i.e.* amplification for the creation of plasmid resources or repair templates for transformation), whereas BioTaq was used for screening amplifications.

PCRs were carried out according to manufacturer's protocols for each of these products. Briefly, Phusion reactions were run in 20 or 50 µL reactions, with an example 50 µL Phusion reaction containing the below components:

| **Phusion PCR Reaction Component** | **50 µL Reaction** |
|---|---|
| Nuclease-free water | To 50 µL |
| 5x Phusion HiFi or GC buffer | 10 µL |
| 10 mM dNTPs | 1 µL |
| 10 µM forward primer | 2.5 µL |
| 10 µM reverse primer | 2.5 µL |

| | |
|---|---|
| Template DNA | Variable : < 250 ng |
| Phusion DNA polymerase | 0.5 µL |

For challenging GC-rich templates, 1.5 µL DMSO was added to the reaction.

Cycling followed the following program:

| **Phusion PCR Cycling Conditions** | **Temperature** | **Time** |
|---|---|---|
| Initial denaturation | 98°C | 30 seconds |
| 25 – 35 cycles | 98°C | 5 – 10 seconds |
| | 45 – 72°C | 10 – 30 seconds |
| | 72°C | 15 – 30 seconds / kb |
| Final extension | 72°C | 5 – 10 minutes |
| Hold | 4°C | ••• |

For reactions carried out with BioTaq the reaction was set up following the below protocol:

| **BioTaq PCR Reaction Component** | **50 µL Reaction** |
|---|---|
| Nuclease-free water | To 50 µL |
| 10x $NH_4$ Reaction Buffer | 5 µL |
| 50 mM $MgCl_2$ Solution | 3 µL |
| 10 mM dNTPs | 5 µL |
| 10 µM forward primer | 2.5 µL |
| 10 µM reverse primer | 2.5 µL |
| Template DNA | Variable : < 250 ng |
| BIOTAQ | 1 µL |

Cycling followed the below program:

| **BioTaq PCR Cycling Conditions** | **Temperature** | **Time** |
|---|---|---|
| Initial denaturation | 94°C | 30 seconds |
| 25 – 35 cycles | 94°C | 5 – 10 seconds |
| | 45 – 72°C | 10 – 30 seconds |
| | 72°C | 15 – 30 seconds / kb |
| Final extension | 72°C | 5 – 10 minutes |
| Hold | 4°C | ••• |

In both cases, the PCR cycling programs were run in a BioRad DNA Engine Thermal Cycler. Annealing temperatures were calculated with the NEB Tm calculator tool [19].

## 2. PCR cleanup

PCR amplicons were cleaned for manipulations such as ligation using a Sigma Aldrich GenElute PCR Cleanup Kit (Sigma, Cat #NA1020-1KT) according to manufacturer guidelines.

## 3. Restriction enzyme digest of DNA

DNA digests were performed using NEB restriction enzymes following manufacturer guidelines.

## 4. Agarose gel electrophoresis

Gel electrophoresis was used to identify the presence of DNA of a certain size, in particular following PCR amplification reactions or restriction enzyme digest. Gels were typically made up with 1.5% agarose (Lonza SeaKem LE Agarose, Cat #50004) in 1x **TBE buffer**, with higher concentrations used for DNA fragments shorter than 300 base pairs and lower percent gels for DNA fragments in a range from 1,500 – 10,000 base pairs. Gels included RedSafe (ChemBio, Cat #21141) for DNA imaging. DNA was mixed with **Orange Loading Dye** for loading into the wells. Gels were run in an electrophoresis tank with 1x **TBE** at 90 – 120 V until the DNA was separated and imaged on a UVP BioDoc-It Imaging System UV transilluminator.

## 5. Gel DNA extraction

To purify DNA from an agarose gel, the gel was viewed by UV on a transilluminator and sections containing the DNA of desired size cut out with a sterile scalpel. The DNA from these gel fractions was purified using a QIAquick gel extraction kit (Qiagen, Cat #28704) following manufacturer guidelines.

## 6. Golden gate cloning

To create plasmid constructs for yeast transformation, golden gate cloning using parts and the design strategy from the Yeast Toolkit [20] were employed. Briefly, plasmids were created or PCR amplifications were performed such that the DNA fragment of interest was flanked by BsaI restriction enzyme cut sites. Upon BsaI (NEB, Cat #R0535S) digestion, the plasmids or amplicons including the DNA fragments of interest were cut, leaving four-base overhangs designed to allow specific directional ligation of the parts. The digestion-ligation assembly reactions were set up as follows:

| Component | Volume |
|---|---|
| T4 DNA ligase buffer (NEB, Cat #M0202S) | 1 µL |
| T7 DNA ligase (NEB, Cat #M0318S) | 0.5 µL |
| BsaI (NEB, Cat #R0535S) | 0.5 µL |
| Parts Plasmids | 20 fmol each part, concentrated by Rotovap to < 8 µL |
| Water | To 10 µL |

The reaction mixtures were cycled in a thermocycler following the program:

| | Temperature | Time |
|---|---|---|
| Initial denaturation | 98°C | 30 seconds |
| ≥ 25 cycles | 42°C | 2 minutes |
| | 16°C | 5 minutes |
| Final digestion | 60°C | 10 minutes |
| Heat inactivation | 80°C | 10 minutes |
| Hold | 4°C | ••• |

## 7. DNA quantification

DNA concentration and purity were measured on a DeNovix DS-11 FX spectrophotometer and fluorimeter. The concentration of double stranded DNA, particularly in creating libraries for NGS, was measured with Qubit fluorimetry (Invitrogen, Cat #Q32857) using the High Sensitivity DNA protocol.

## 8. DNA fragment size determination

DNA fragment length was measured using a High Sensitivity (HSD1000) ScreenTape (Agilent, Cat #5067- 5584) with High Sensitivity D1000 reagents (Agilent, Cat #5067- 5585) on an Agilent 4200 TapeStation following manufacturer's instructions. The Agilent TapeStation Analysis software was used to calculate the peak value between 150 and 1400 base pairs for sequencing library fragment length determination.

## 9. Sequencing library quantification

DNA sequencing library concentrations were analyzed by qPCR using the NEBNext Library Quant kit (NEB, Cat #E7630L). This was run following manufacturer guidelines. Briefly, DNA libraries were diluted 1:10,000 or 1:100,000 using the NEB dilution buffer. 4 µL of diluted library was added to 16 µL of qPCR master mix supplemented with the supplied primer mix. Each sample was run in triplicate, as were the DNA standards and the no template control. A RotorGene Q (Qiagen) qPCR machine was used to run the cycling program (95°C for 60 seconds, 35 x [95°C for 15 seconds, 63°C for 45 seconds]) and to quantify fluorescence. The Q-Rex (Qiagen) software was used to generate Ct values. The Ct values of the four known standards were used to generate a standard curve, from which the concentration of the experimental samples was inferred. This was used in conjunction with average library fragment size information measured by TapeStation to inform library pooling and loading for sequencing runs.

## 10. Repli-Seq

Extracted DNA was sonicated to an average fragment length of 300 bases on a Bioruptor UCD-200 diagenode. DNA was purified using SPRI beads as per

supplier protocol and end-repaired with the NEBNext End Prep Kit (NEB, Cat #E7645L). Sequencing adapters were ligated on using the NEBNext Ultra II ligation kit. DNA was purified using AMPure XP beads (Beckman Coulter, Cat #B79194-AB). BrdU pulldown was carried out as per (Peace *et al.*, 2016) [21]. Briefly, DNA was heat-denatured and snap-cooled, then incubated overnight with anti-BrdU antibody (BD Biosciences, Cat #BD347580). The sample was then incubated for one hour with Protein G dynabeads (Thermofisher, Cat #10003D), washed three times in **IP Buffer** and once with **Tris-EDTA** (TE). The DNA was eluted with TE/1% (w/v) sodium dodecyl sulfate (SDS) at 65°C. Eluted DNA was cleaned using AMPure XP beads, and then amplified with Illumina indexing primers using 16 cycles of the NEBNext Ultra II kit (Index Primers Set 1: Cat #E7335L and Set 2: Cat #E7500L). Amplified libraries were cleaned, quantified, checked for fragment length, and pooled for sequencing as described above.

Libraries were sequenced using 75 cycles on an Illumina NextSeq 500 for a minimum of 10 million reads per library.

### 11. Synchronized spike-in ChIP sequencing

Starter cultures of the experimental *S. cerevisiae* strains of interest and *S. bayanus* spike-in strain were inoculated into desired medias. *S. cerevisiae* cells were inoculated from the starter culture into 100 mL of media to achieve an OD of 0.1. This culture was grown for one doubling, at which point 2 mL of the culture was harvested for flow cytometry and 2 mL for Western blotting. 15 µL of 3 mM alpha factor (BioServ, custom synthesized peptide: WHWLQLKPGQPMY) was added per 100 mL of cell culture and the cells were incubated for at least 1.5 doubling times, with additional alpha factor added every 1/3 of a doubling time.

The cells were assessed by microscopy, and arrest was deemed complete when over 95% of the cells were unbudded.

Upon alpha factor arrest, 2 mL of culture was harvested for flow cytometry and 2 mL for Western blotting. Cells were then released from G1 arrest with the addition of pronase (Calbiochem, Cat #53702) and either allowed to progress through S phase or arrested in early S phase with the addition of HU (Sigma, Cat #H8627) to a final concentration of 200 mM as specified in the experiment. At this point, culture aliquots were also left non-induced or induced by the addition of galactose (Sigma, Cat #G0625) as specified in the experiment. Further samples taken for flow cytometry used 0.5 mL of culture.

At the specified timepoint for ChIP-Seq harvest, cell density of the *S. cerevisiae* and *S. bayanus* cultures was measured by spectrophotometry ($OD_{600}$). The volume of *S. bayanus* culture required for a 5% spike-in for each *S. cerevisiae* sample was calculated and that volume added to each of the samples. Each yeast sample culture was then fixed with 37% formaldehyde (1.4 mL / 50 mL culture) (Sigma, Cat #252549). This was incubated at room temperature for 15 minutes with occasional shaking. The cross-linking was quenched by addition of 2.5 M glycine (7 mL / 50 mL culture) and incubation at room temperature for 5 minutes with occasional agitation. The cultures were then transferred into 50 mL falcon tubes and spun at 5000 g for 5 minutes at 4°C. The cells were washed twice in 40 mL ice-cold **TBS**, resuspended in a small volume (<1 mL) of ice-cold **TBS**, and transferred to an Eppendorf tube. The Eppendorf tubes were spun for 2 minutes at 4°C and as much supernatant as possible was removed from the pellet. The pellet was snap frozen and stored at -80°C for up to a month.

When proceeding to cell lysis and chromatin harvest, the cell pellets were lysed by adding 600 µL of glass beads to a skirted, screw-capped FastPrep tube on ice. The cell pellet was thawed from -80°C on ice, and then resuspended in 600 µL **lysis buffer**. The volume was transferred to the FastPrep tubes and placed in the FastPrep-24 bead beater (MP Biomedicals, Cat #116004500). The cells were beat following a protocol of four repeats of max speed for 20 seconds followed by incubation on ice for 2 minutes. At the end of the FastPrep protocol, a hole was punched in the bottom of the FastPrep tubes with a needle (26G, 0.45x13 mm) and the tubes were spun in a 15 mL falcon tube for 1 minute at 900 g and 4°C. The volume was collected and transferred to an Eppendorf tube. It was then subjected to sonication by the tip sonicator following a protocol of five repeats of 15 seconds of sonication followed by a 1 minute spin (21k g, 4°C). If bubbles formed during the 15 second sonication, the spin was increased to 3 minutes. Following the final sonication, the tube was spun for 5 minutes (21k g, 4°C) to pellet cell debris. The supernatant was transferred to a new Eppendorf tube, and spun for 2 minutes (21k g, 4°C). The maximum possible supernatant was transferred to a new Eppendorf labeled ‘WCE.’ 2 µL of protease inhibitor cocktail (cOmplete ULTRA, Roche/Sigma, Cat #11873580001) was added, and the tube was stored at -80°C for up to a month if required.

The magnetic beads for the immunoprecipitation were prepared by taking 50 µL of protein G dynabeads (Invitrogen, Cat #10004D) per sample and per IP to a 15 mL falcon. The beads were spun at 3000 g for 1 minute, then resuspended in 10 mL **PBS/BSA** solution. This was repeated for a second wash, after which the beads were resuspended in a total volume of **PBS/BSA** to allow for 250 µL / sample / IP. The bead solution was split into the total number of distinct immunoprecipitations

to be carried out (*i.e.* split into two aliquots of 500 µL each if two samples would be subjected to both Flag and Myc immunoprecipitations). The appropriate antibody was added to the beads to allow for 6 µL per sample [Antibody Table]. This solution was incubated overnight at 4°C on a rotating wheel.

Following overnight incubation, the beads were spun for 1 minute at 2000 g and 4°C. The beads were washed five times with 2 mL of ice-cold **PBS/BSA**. The supernatant was completely removed, and the beads were resuspended in a total volume of **PBS/BSA** to correspond to 30 µL / sample and aliquoted into one Eppendorf tube for each sample and each IP. The chromatin was thawed on ice from -80°C and 200 µL of chromatin was added to an appropriate antibody tube. The tubes with antibody-conjugated beads and chromatin were incubated overnight on a rotating wheel at 4°C.

After overnight incubation, the beads were pelleted on a magnetic stand in the cold room. The supernatant was removed and beads were resuspended in 1 mL of ice-cold **lysis buffer**. The tubes were incubated for 1 minute on a rotating wheel at 4°C, pulse spun, and then pelleted on the magnetic stand. The wash with **lysis buffer** was repeated. The beads were then resuspended in 1 mL **lysis buffer** supplemented with 360 mM NaCl. This was incubated for 1 minute on a rotating wheel at 4°C and then the beads pelleted on a magnetic stand and supernatant removed. The beads were then washed (following above procedure) twice with **wash buffer** and once with **TE** (pH 8). The beads were pelleted by spinning for 2 minutes at 21k g and 4°C. All supernatant was carefully removed without allowing the beads to dry. 100 µL **TE-SDS** solution was added to the beads and the tube was labeled 'IP.' In parallel, the remaining WCE for each

sample was thawed from -80°C and 5 µL was taken to a new tube labeled ‘Input’ and to it added 95µL **TE-SDS**. Both sets of input and IP tubes were vortexed for 1 minute and left shaking at 65°C for 9 – 15 hours to complete reverse crosslinking.

DNA purification was carried out by spinning the tubes for 2 minutes at 21k g. The supernatant was transferred to a new Eppendorf and 5 µL of RNAse A (10 mg/mL, Sigma Cat #R6513) was added. The tubes were incubated at 37°C for 30 minutes. 10 µL Proteinase K (20 mg/mL, Sigma, Cat #P6556) was added. The tubes were left incubating at 37°C overnight. 200 µL of **TE** (pH 8) was added to the tubes, followed by 300 µL phenol-chloroform (25:25:1, pH 8.0, Sigma, Cat #77617-500ML). The tubes were vortexed for 30 seconds, and then spun for 5 minutes at 21k g. The aqueous phase was transferred to a new tube, and a second phenol chloroform extraction was carried out. A third extraction was carried out with the addition of 600 µL chloroform. The aqueous phase was transferred to a new tube, and 0.1 volume (20 µL) of 3 M NaOAc (pH 5.2) and 1 volume (200 µL) of isopropanol were added. The tube was mixed by inversion, left in the freezer for 30 minutes, and then spun for 40 minutes at 21k g and 4°C. The resulting pellet was washed with 1 mL ice-cold 70% ethanol, and then resuspended in 50 µL milli-Q $H_2O$ and stored at -20°C. This DNA was used as starting material for DNA sequencing library creation using the NEBNext Ultra II DNA library prep kit (NEB, Cat #E7370L) and following the NEB protocol. The input DNA concentration was in a range from 5 – 100 ng, depending on IP and DNA extraction yield.

## 12. Chromatin fractionation

Cells were grown to $OD_{600}$ 0.4 in 25 mL **YPAD**. After desired treatment, cells were transferred to a 50 mL falcon tube with 20 mL of ice (milli-Q $H_2O$), 450 µL of 10% sodium azide, and 225 µL of 200 mM phenylmethylsulfonyl fluoride (PMSF). This was spun down (3000 g, 5 min, room temperature) and the cell pellet was resuspended in 200 µL ice-cold **wash buffer A** and transferred to an Eppendorf tube.

The cells were spun at 2000 g for 3 minutes at 4°C, and the supernatant was carefully removed. The cell pellet was weighed to determine the amount of zymolyase (T20, 5 mg/mL, Seikagaku Biobusiness, Cat #120491) required for the **spheroplasting buffer** (0.4 mg zymolyase / 100 mg wet cell pellet). The pellet was loosened by tapping the bottom of the tube and then resuspended in 200 µL of **spheroplasting buffer** and incubated at 30°C for 10 minutes. The spheroplasting was checked under the microscope by mixing 5 µL of the cell suspension with 1 µL of 10% Triton X-100. The spheroplasting was deemed complete at the point when most of the cells burst after 30 seconds in Triton X-100. If spheroplasting was not complete, the incubation was extended and continued checks for spheroplasting were carried out every 3 minutes. Following spheroplasting, the tubes were put on ice and then spun down at 2000 g and 4°C for 3 minutes. The supernatant was removed and the pellet resuspended in 200 µL of 0.4 M **sorbitol/EB**. The suspension was spun down again at 2000 g and 4°C for 3 minutes and the supernatant removed. The pellet was again weighed and its volume estimated taking 100 mg as 100 µL. The pellet was resuspended in **buffer EB** to reach a total volume of 200 µL. 5 µL of Triton X-100 (Sigma, Cat #X100) was added and the tube was gently mixed by inversion and then

incubated on ice for 5 minutes. 150 µL of the lysate was aliquoted onto the top of a 500 µL **sucrose cushion**. The tube was spun at 13.5 g for 10 minutes at 4°C. Meanwhile, 45 µL of the residual lysate was moved to a new tube with 15 µL **4x LDS loading buffer**. This was labeled 'WCE' and boiled for 5 minutes.

When the spin of the lysate on the sucrose cushion was complete, 75 µL of the supernatant was moved to a new tube and mixed with 25 µL **4x LDS loading buffer**. This was labelled 'SN1' and boiled for 5 minutes. The remaining supernatant was aspirated away and the pellet was washed with 500 µL **EBX**. The tube was spun at 3000 g for 3 minutes at 4°C and the pellet resuspended in 45 µL **EBX**. 1 µL of benzonase (Sigma, Cat # E8263-5KU) was added, and the reaction was incubated on ice for 30 minutes. 15 µL of **4x LDS loading buffer** was added to this tube and it was boiled for 5 minutes and then spun at 3000 g for 3 minutes at room temperature. The supernatant was transferred to a new tube and labelled 'P1.'

**13. Western blotting**

**Acrylamide gels** (7% running, 6% stacking) were poured. Samples were run in 1x **running buffer** at 120 V. Proteins were transferred onto a 0.45 µm PVDF membrane (Amersham Hybond-P) in **transfer buffer** at 30 V and 0.09 A in a Bio-Rad Mini-PROTEAN II Cell apparatus. The transfer was run overnight at 4°C. Total protein was visualized with **Ponceau** staining, following which the membrane was blocked for 30 minutes with 5% milk (Marvel Original Skimmed Milk Powder) in **PBST**. Primary staining was carried out for 2 hours at room temperature as per the Antibody Table unless otherwise noted. This was followed

by three washes in **blocking buffer**, and then secondary staining for 45 minutes at room temperature, with antibodies as specified [Antibody Table].

For blots imaged with chemiluminescence (*i.e.* an HRP-conjugated secondary antibody), following secondary antibody incubation the membrane was washed four times for 5 minutes each with **PBST**. ECL solution (Amersham, Cat #RPN2232SK) was made, pipetted onto the membrane, and left to incubate for 1 minute. The blot was then imaged between two sheets of plastic using a LAS4000 luminescence imager.

For blots imaged with fluorescent-conjugated secondary antibodies, following secondary antibody incubation the membrane was washed three times for 5 minutes each with **PBST**, and then one time with PBS. The blots were imaged with an Oddysey CLx scanner and analyzed using ImageStudioLite.

**Antibody Table.** Antibody usage for Western blotting, immunofluorescence, flow cytometry, Repli-Seq, and ChIP.

| Antibody | Catalogue Number | Usage | Dilution |
|---|---|---|---|
| Mouse anti-Flag | Sigma M2 F1804 monoclonal | Western | 1:1,000 (5% (*w/v)* milk / PBST) |
| | | ChIP | 6 µL / IP |
| Rabbit anti-Myc | ProteinTech 16286-1-AP polyclonal | Western | 1:1,000 (5% (*w/v)* milk / PBST) |
| | | ChIP | 6 µL / IP |
| Rat anti-Tubulin | Abcam ab6161 monoclonal | Western | 1:1,000 (5% (*w/v)* milk / PBST) |
| Rabbit anti-RAD21 (SCC1) | Abcam ab154769 polyclonal | IF | 1:500 (5% (*w/v)* BSA / PBS) |

| | | | |
|---|---|---|---|
| Mouse anti-BrdU | BD Biosciences BD347580<br><br>monoclonal | Flow | 1:200 (flow wash solution) |
| | | IF | 1:200 (5% (*w/v)* BSA / PBS) |
| | | Repli-Seq | 15 µL / IP |
| Goat anti-Rabbit IRDye 680 | LiCOR P/N 925-68071 | Western | 1:10,000 (1% (*w/v)* milk / PBST) |
| Goat anti-Rat IRDye 680 | LiCOR P/N 926-68076 | Western | 1:10,000 (1% (*w/v)* milk / PBST) |
| Goat anti-Mouse AlexaFluor 647 | Invitrogen A32728 | Flow | 1:100 (flow wash solution) |
| | | Imagestream | 1:500 (flow wash solution) |
| | | Western | 1:10,000 (1% (*w/v)* milk / PBST) |
| Goat anti-Rabbit Alexafluor 647 | Invitrogen A21245 | IF | 1:1,000 (5% (*w/v)* BSA) |
| Rabbit anti-Mouse AlexaFluor 568 | Invitrogen A11061 | IF | 1:1,000 (5% (*w/v)* BSA) |
| Goat anti-Rat HRP | Jackson Laboratories 112-035-167 | Western | 1:10,000 (1% (*w/v)* milk / PBST) |
| Goat anti-Mouse HRP | Jackson Laboratories 115-035-062 | Western | 1:10,000 (1% (*w/v)* milk / PBST) |
| Goat anti-Rabbit HRP | Jackson Laboratories 111-035-144 | Western | 1:10,000 (1% (*w/v)* milk / PBST) |

Bacterial culture and manipulation

**1. Growth**

*E. coli* was grown at 37°C, with liquid cultures in a shaking incubator. Luria-Bertani (LB) agar plates and LB broth for *E. coli* growth were sourced from the Dunn School of Pathology media and stores facility.

### *Ampicillin selection for E. coli*

An ampicillin stock at 100 mg/mL was made up from ampicillin sodium salt (Sigma, Cat #9518) and filter sterilized (0.45 µm). Cooled autoclaved LB agar or broth was supplemented with ampicillin stock solution to a final concentration of 100 µg/mL.

### *Chloramphenicol selection for E. coli*

A chloramphenicol stock at 25 mg/mL in ethanol was made up from chloramphenicol (Sigma, Cat #C0378). Cooled autoclaved LB agar or broth was supplemented with chloramphenicol stock solution to a final concentration of 25 µg/mL.

## 2. Heat shock

Heat shock competent DH5α cells were incubated in Eppendorf tubes on ice with the transforming DNA for a minimum of 20 minutes. Cells were exposed to 42°C for 45 seconds and then returned to ice for 2 minutes of recovery. 300 µL LB media was added to cells, and the culture was incubated at 37°C for 30 – 60 minutes. The culture was then pipetted onto LB agar plates with appropriate antibody for selection, and the culture spread with sterile glass beads. The plates were incubated at 37°C for 24 hours, at which point colonies could be selected for screening.

### 3. Colony PCR

Initial screening of bacterial colonies was carried out using colony PCR. Briefly, BioTaq master mix was made up according to the number of colonies and the number of primer sets intended for screening. One PCR tube was aliquoted with 39.5 µL milli-Q water for each primer set for each colony (*e.g.* for screening 6 colonies with 2 primer sets, 79 µL water would be aliquoted to each of 6 PCR tubes). A small amount of colony was picked from an agar plate and resuspended in the water with a pipette tip. The water with bacterial suspension was aliquoted for each PCR reaction for each colony, and 10.5 µL of PCR master mix was added to each bacterial suspension aliquot. The reaction was then cycled as appropriate for the primers and amplicon and products assessed by agarose gel electrophoresis.

### 4. Plasmid miniprep

*E. coli* cells harboring the plasmid of interest were grown at 37°C shaking in LB liquid culture with a selective antibiotic for 12 – 16 hours. The cells were collected by centrifugation (3000 g, 5 minutes, room temperature) and the pellet was frozen at -20°C or brought forward immediately for plasmid harvest. A NucleoSpin plasmid kit (Machery-Nagel, Cat #740588.250) was used according to manufacturer's instructions.

## Yeast culture and manipulation

Strains used are listed in the Strain Table. Yeast strains used were of W303 background [22] unless otherwise noted.

## 1. Growth

Media was made up according to manufacturer's protocols and autoclaved at 117ºC for 15 minutes. *S. cerevisiae* strains were grown at 30°C unless otherwise noted, for example for those strains including a *cdc7* temperature sensitive mutation which were grown at 23°C. *S. bayanus* strains were grown at 23°C.

### *Yeast-Peptone-Adenine-Dextrose (YPAD) medium*

Formedium CCM1010 (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose, 0.04 g/L adenine sulphate)

### *Yeast-Peptone-Adenine-Dextrose (YPAD) agar*

Formedium CCM0510 (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose, 0.04 g/L adenine sulphate, 20 g/L agar)

### *YPA*

(Bacto Yeast extract BD 212750 10 g/L, Bacto Peptone BD 211820 20 g/L, 0.5% adenine solution). This was adjusted to pH 6 – 6.5 with 1 M HCl.

### *Sporulation plates*

Potassium acetate plates were used for yeast sporulation. Potassium acetate (Sigma, Cat #236497, 20 g/L), Bacto Yeast Extract (BD, Cat #212750, 2.2 g/L), Dextrose (Sigma, Cat #G8270, 0.5 g/L), COM drop out powder (Formedium, Cat #DCS0031, 0.87 g/L), agar (Fluka, Cat #05039, 25 g/L). This was adjusted to pH 7.0 with 1 M HCl.

### *G418 selection medium*

G418 disulfate salt (Sigma, Cat #A1720) was used to make a stock solution at 100 mg/mL. This stock was filter sterilized with a 0.45 µm filter. The stock was

added to cooled autoclaved YPAD agar or broth for a final concentration of 400 µg/mL.

### *CloNat selection medium*

A 100 mg/mL stock of CloNat (Werner BioAgents, Nourseothricin dihydrogen sulfate) was made up in water and filter sterilized. This stock was added to cooled autoclaved YPAD agar or broth for a final concentration of concentration of 100 µg/mL.

### *Raffinose*

A 20% raffinose stock was made up of D (+) raffinose pentahydrate (Sigma, Cat #83400) dissolved in YPA media and filter sterilized. This was stored at room temperature and added to media for a final concentration of 2% unless otherwise stated.

### *Galactose*

A 20% galactose stock was made up of D (+) galactose (Sigma, Cat #G0625) dissolved in YPA media and filter sterilized. This was stored at room temperature and added to media for galactose concentrations between 0.01 and 2%.

### *Estradiol*

A 10 µM β estradiol (Sigma, Cat #E2758) stock was made up in ethanol and stored at -20°C. This was added to yeast media for a final estradiol concentration of 1 µM.

## 2. Transformation

*S. cerevisiae* strains were transformed following a protocol published by Gietz and Schiestl [23]. Briefly, cells were grown to $OD_{600}$ 0.5 – 1.0. Cells were spun down

(3000 g, 5 min, room temperature) and washed twice with 10 mL milli-Q water, and then resuspended in 1 mL milli-Q water and transferred to a microcentrifuge Eppendorf tube and spun (13.5k g, 30 seconds). The pellet was resuspended in 1 mL milli-Q water and 100 µL aliquots taken to a fresh tube for each transformation. The aliquots were pelleted and supernatant removed. The pellet was resuspended in 360 µL transformation mix comprised of: 240 µL PEG 3350 (50% w/v), 36 µL 1 M lithium acetate, 50 µL 2 mg/mL salmon sperm carrier DNA, and 34 µL plasmid or linearized DNA (up to 30 µg) made up in milli-Q water. *S. cerevisiae* cells were then heat shocked at 42ºC for 40 minutes. *S. bayanus* cells were heat shocked by incubation at 37ºC for 40 minutes. Heat-shocked cells were pelleted (13.5k g, 1 min) and recovered in liquid YPD before being plated onto selective media, or alternatively were directly plated onto YPD and replica plated onto selective plates after overnight growth.

**3. Yeast mating**

Yeast were mated by plating a small amount of a single colony from a mating type α and spreading overtop of it a similar amount from a single colony of the mating type a strain. The plate was left overnight at a temperature supporting grown (30°C in all cases except for yeast with the *cdc7-1* temperature sensitive mutation, which were left at 23°C). The plate was then streaked to single colonies on YPAD.

**4. Yeast sporulation**

Mated yeast were sporulated by streaking onto high potassium acetate (KAc) plates. The plates were left at room temperature for 3 – 7 days until spores had formed, as checked by microscopy.

**5. Tetrad dissection**

1 µL of zymolyase stock (T20, 5 mg/mL, Seikagaku Biobusiness, Cat #120491) was added to 19 µL **dissection buffer** in an Eppendorf tube. A small volume of sporulated tetrad yeast from the sporulation plate was added to this buffer, and the Eppendorf was left at 37°C for 15 – 30 minutes (optimized for strain).

A flat YPAD dissection plate was cut with two grooves along the diameter, leaving a central lane, and was left open in a laminar flow hood to dry. After the desired incubation time allowing for sufficient zymolyase digestion, the dissection buffer yeast mixture was diluted with 40 – 200 µL $H_2O$ and spread along the channel between the two grooves. The plate was left open in the hood for 10 – 15 minutes to dry. The plate was placed upside down on the dissection microscope (Singer, SporePlay) and the glass needle was used to isolate tetrads and separate the four spores across a row on the plate. This plate with dissected tetrads was then left to grow until colonies could be seen, at which point replica plating and PCR were used to identify the desired strains.

### 6. Freezing down strains for long-term storage

Strains desired for freezing down were plated onto ¼ to ½ of a YPAD plate and grown overnight. A pipette tip was used to scrape the plate-grown culture into a vial with approximately 1 mL of autoclaved 25% glycerol. The vial was vortexed and stored at -80°C.

### 7. Yeast synchronization

The yeast pheromone alpha factor was used for cell cycle arrest of *MATa*, *BAR1* wildtype *S. cerevisiae* in late G1. Cells were grown to just below $OD_{600}$ 0.2 and 15 µL of alpha factor at 3 mM (BioServ, custom synthesized peptide: WHWLQLKPGQPMY) was added for every 100 mL of cells. Cells were incubated

for at least 1.5 doubling times, with intermittent addition of alpha factor as above every 30 minutes. Cell cycle arrest was monitored by microscopy and arrest was deemed satisfactory when over 95% of cells were unbudded but forming schmoos. Cells were released with either filter washing or the addition of pronase (200 µg/mL, Calbiochem, Cat #53702) and allowed to proceed through S phase. 500 µL or 1 mL samples were taken before arrest, at the point of full arrest, and following release to confirm cell cycle stage via staining for DNA content and flow cytometry.

Hydroxyurea (HU, Sigma, Cat #H8627) was used to arrest cells in early S phase via inhibition of replication fork elongation. Generally, HU arrest was used upon release from alpha factor G1 arrest. Alpha factor arrested cells were synchronously released via pronase or filter washing into media with 200 mM HU.

**8. Flow cytometry**

At the point of harvesting cells for flow cytometry, cells were fixed with 1/100$^{th}$ volume 10% sodium azide, vortexed, and stored on ice or at -20°C until the full set of samples was ready for processing. At that point, samples were spun at 15k g for 30 seconds. The samples were rinsed with 1 mL milli-Q water, pelleted, and resuspended in 1 mL 70% ethanol. Samples were left in 70% ethanol at 4°C overnight for fixation.

Fixed cells were washed twice in 1 mL sodium citrate (50 mM, 0.45 µm filter sterilized, pH 7.0). Cells were pelleted by centrifugation (15k g, 1 minute, room temperature). Cells were then resuspended in 1 mL 50 mM sodium citrate with RNase A (0.25 mg/mL). The cells were incubated with RNase A for 1 hour at 30°C, after which 20 µL of 20 mg/mL proteinase K was added to the tube. The

cells were incubated at 50°C for an additional hour. Samples were spun (15k g, 1 minute, room temperature) and then resuspended in 1 mL 50 mM sodium citrate with 1x SYTOX green (Life Technologies Ltd, Cat #S7020). The samples with SYTOX were kept protected from light and incubated either at 4°C for at least 12 hours or at room temperature with periodic agitation for at least 2 hours. Samples were sonicated with the tip sonicator for three pulses of 6 seconds before analysis on a flow cytometer.

Flow cytometry was performed on a BD LSR Fortessa X-20 cell analyzer (BD Biosciences) with detection of forward scatter (FS), side scatter (SS), and BL1 to detect SYTOX green. The parameters were optimized for each experiment but fell within the ranges: FS 580 – 650 V, SS 250 – 305 V, BL1 400 – 430 V, linear acquisition. Events were gated for cells (gating out cellular debris) and singlets (gating out doublets).

**9. Growth curve**

Cell doubling times were measured by inoculating cells into liquid culture and growing to log phase. Culture density ($OD_{600}$) readings were taken over a period spanning at least two doublings within the exponential growth phase. Culture samples taken for OD measurement were diluted as necessary to keep the OD readings within the attenuance linear range from 0.1 – 0.5.

**10. Yeast 'Smash and Grab' DNA extraction**

Yeast cells were harvested by centrifugation and resuspended in one volume of **lysis buffer** in a 2 mL skirted screw cap tube. One volume of phenol-chloroform-isoamyl alcohol (25:25:1, pH 8.0. Sigma, Cat #77617-500ML) was added to the cells in lysis buffer. Approximately 0.3 mL of glass beads were added to the tube,

and the tube was shaken in the FastPrep-24 bead beater at maximum power for 60 seconds, three times. One volume of **TE buffer** was added and the tube mixed by shaking.

The tube was then spun (13.5k g, 5 minutes, room temperature) and the upper aqueous layer transferred into a new tube. Two volumes of chloroform were added, and the tube was shaken vigorously. The tube was spun (13.5k g, 5 minutes, room temperature) and the upper aqueous phase again transferred into a new tube. 10 µL of RNAse A (Sigma R6513, 10 mg/mL) was added and the tube was incubated at 37°C for 10 – 30 minutes, followed by the addition 10 µL of Proteinase K (Sigma P6556, 20 mg/mL) and incubation at 50°C for 1 hour. If the DNA needed to be highly purified, an additional phenol-chloroform-isoamyl extraction was carried out. If not, two volumes of chloroform were added, the sample shaken and spun (13.5k g, 5 minutes, room temperature). The aqueous phase was transferred to a new tube and the DNA was precipitated by adding 0.1 volume of 3 M KoAc (pH 5.5) and 1 sample volume of isopropanol. The tube was left at room temperature for 10 minutes and then spun (max speed, 15 minutes, 4°C). The supernatant was carefully removed and the pellet cleaned with a wash of 1 mL of 70% ethanol before resuspending in **TE** or water.

### 11. Quick SPZ yeast DNA extraction

**SPZ buffer** was aliquoted into PCR tubes with 2 µL per sample. A pipette tip was used to pick up a small volume of yeast cells from a colony to resuspend in the **SPZ buffer**. This was incubated in a PCR machine at 37°C for 30 minutes followed by 95°C for 5 minutes. 30 µL of milli-Q water was added to each sample, and 1 µL of this was used as a template for screening PCRs.

### 12. Yeast total protein extraction

Protein extraction was carried out as per (Kushnirov, 2000)[24]. Approximately 3 x $10^7$ log phase cells growing in liquid media were harvested by centrifugation. The pellet was washed once in water – at this stage it could be frozen if further sample collection was required (for example, in a time-course). The pellet was resuspended in 100 µL milli-Q water and 0.2 M NaOH to a final concentration of 0.1 M. This was incubated at room temperature for 5 minutes, and then pelleted by centrifugation. The pellet was resuspended in 50 µL **sample buffer** and heated to 85 – 100°C for 3 – 5 minutes. The sample was pelleted one final time and the supernatant transferred to a fresh tube, and from there loaded for SDS-PAGE and Western blots.

## Human cell culture and manipulation

### 1. Cell culture

HCT116 cells were grown in MyCoy's 5A media (Sigma, Cat #M9309) with 10% FBS (Gibco, Cat #10500), 100 U/mL penicillin and 100 µg/mL streptomycin (Pen-Strep Gibco, Cat #15140-122) at 37°C with 5% $CO_2$.

### 2. Cohesin degradation

Cells were treated with 500 µM indole-3-acetic acid (IAA, Sigma, Cat #I5148) to induce loss of mAID-tagged SCC1.

### 3. BrdU pulse

100 µM BrdU (Sigma, Cat #B5002-1G) was added to the media of exponentially-growing cells to label newly synthesized DNA. For Repli-Seq, cells were incubated for 2 hours in BrdU. For immunofluorescence and ImageStream analysis of BrdU foci, cells were incubated for 30 minutes in 50 µM BrdU.

## 4. Synchronization

Asynchronously-growing cells were treated with 2.5 mM thymidine for 24 hours and released by washing into fresh media. 3 hours after release, cells were treated with 100 ng/mL nocodazole (Calbiochem, Cat #487928) for 8 hours, at which point mitotic cells were collected via shake-off. Cells were plated into fresh media without nocodazole. For BrdU-IP experiments with cohesin depletion, 30 minutes after replating from the mitotic shake-off, cells were treated with 500 μM IAA or left untreated. Cells were collected at 4, 7, or 10 hours post release from nocodazole, and BrdU was added to 100 μM 2 hours before the collection of each cell population (*i.e.* at 2, 5, and 8 hours).

## 5. Fluorescence activated cell sorting (FACS)

Cell sorting was carried out as described in (Marchal *et al.*, 2018)[25]. Briefly, asynchronously-growing cells were washed, trypsinized, and harvested, and then fixed with 75% ethanol. Fixed cells were prepared for fluorescence activated cell sorting (FACS) with 250 μg/mL RNase A and 50 μg/mL propidium iodide (PI, Sigma, Cat #P4170). Cells were sorted on the basis of PI stain into four bins spanning S phase to collect a minimum of 2.5 x $10^5$ cells / bin.

## 6. Flow cytometry

To confirm the loss of the mClover-tagged SCC1 upon auxin treatment, cells were fixed with 75% ethanol and prepared for flow cytometry assessment of mClover signal on a BD LSR Fortessa.

To confirm the cell cycle staging of BrdU-pulsed samples, cells were fixed and permeabilized with 75% ethanol. DNA was denatured via incubation in 0.2 mg/mL pepsin (Sigma, Cat #P7012-1G) in 2 M HCl for 20 minutes. Cells were washed

twice with **wash solution**. Cells were then incubated in **wash solution** with anti-BrdU antibody (BD Biosciences, Cat #BD347580) at 1:200 for 1 hour [Antibody Table]. Cells were washed twice, then incubated with AlexaFluor 647 Goat anti-Mouse (Invitrogen, Cat #A32728) antibody at 1:100 in **wash solution** for 1 hour, followed by two washes. Cells were finally treated with RNase A (250 µg/mL) and PI (50 µg/mL) before acquisition on a BD LSR Fortessa.

### 7. DNA extraction

DNA was extracted using phenol:chloroform followed by an isopropanol precipitation. Plated cells were washed with ice-cold tissue culture grade PBS, then harvested by trypsinization and spun for 5 minutes at 500 g. Supernatant was removed, and the cell pellet was resuspended in 250 µL PBS. 250 µL **2x digestion buffer** was added. This was incubated for 30 minutes at room temperature. 25 µL 20 mg/mL proteinase K was added and the tube was incubated shaking overnight at 65°C. An equal volume of phenol:chloroform was added to the tube, which was then vortexed and spun for 5 minutes at 1700 g. The top aqueous layer was transferred to a new tube. A second phenol:chloroform extraction was carried out as before. After two phenol:extractions, an equal volume of chloroform was added to the aqueous layer in the new tube. This was vortexed and spun, and the aqueous layer transferred to a new tube. 1/10$^{th}$ volume of 7.5 M ammonium acetate and one volume of cold isopropanol were added. This was spun for 1 hour at 4°C and 1700 g. The supernatant was carefully removed, and the pellet was gently washed with 70% ethanol and then air dried. The pellet was finally resuspended in milli-Q water or **TE**.

## 8. Immunofluorescence

### *BrdU replication foci*

Cells were seeded onto coverslips (thickness #1.5) and grown to 80% confluency. Media was replaced and cells were treated with IAA (500 μM) or left untreated for 3.5 hours at the end of which all cells were pulsed with BrdU (50 μM) for 30 minutes.

Upon completion of the BrdU pulse, coverslips were washed with PBS and fixed for 10 minutes at room temperature with 3% (*w/v*) formaldehyde/PBS. Coverslips were washed three times with PBS. After the final wash was removed, cells were permeabilized through incubation in 0.2% (*v/v*) Triton X-100 in PBS at room temperature for 10 minutes. Cells were washed five times with **PBST** and then DNA was denatured with 4 M HCl for 10 minutes. The coverslips were washed and blocked with 5% (*w/v*) BSA (Sigma, Cat #A7906) in **PBST**. Cells were incubated for 1 hour with anti BrdU antibody (BD247580) 1:200 in **IF block solution**. After three washes with **PBST**, coverslips were incubated for 1 hour in the dark with AlexaFluor Rabbit anti-Mouse 568 antibody (Invitrogen, Cat #A11061) at 1:1000 in **IF block solution**. Coverslips were washed with **PBST** and nuclei were stained with DAPI. Coverslips were finally mounted with VectaSheild H1000 (Vector Laboratories, Cat #H1000). DAPI signal was used to identify cells for image acquisition and BrdU-positive cells were classified by eye into five S phase stages while blinded to the treatment status of the images.

### *SCC1*

Asynchronous cultures of HCT116-SCC1-mAID cells were seeded onto coverslips and treated for 2 hours with IAA (500 μM) or left as controls. At the end of

treatment, coverslips were washed with PBS then fixed with 3% formaldehyde. Cells were permeabilized with 0.2% Triton X-100 in PBS, washed with **PBST**, and blocked for 1 hour in **IF block solution**. Coverslips were incubated for 1 hour in anti-SCC1 antibody (Abcam, Cat #154769) at 1:500 in **IF block solution**. After three washes, coverslips were incubated for 1 hour in AlexaFluor 647 Goat anti-Rabbit antibody (Invitrogen, Cat #A21245) at 1:1000 in **IF block solution**. Coverslips were washed, stained with DAPI, and mounted in VectaSheild.

### 9. ImageStream

For visualizing replication foci in synchronized cells, cells were synchronized and pulsed with BrdU as described above. Cells were harvested and prepared for flow cytometery with PI and BrdU staining at lower secondary antibody concentrations than those used for Fortessa flow cytometry (1:500 for anti-Mouse AlexaFluor 647). The cells were analyzed on the ImageStream (Luminex) which acquired cell microscopy images along with fluorescent signal readings.

The ImageStream machine learning image analysis software suite was used for high-throughput identification of cells fitting BrdU patterns characteristic of early, mid, or late S phase. Briefly, images from the full set of samples (auxin-treated and control cells at 4, 7, and 10 hours into S phase) were considered in a single pool. A set of between 20 and 40 cell images representative of each of the S phase patterns (early S - numerous small BrdU foci across the nucleus; mid S - a peripheral ring; late S - large foci) were selected as classifiers. The ImageStream software was run to identify all images from the pooled set that fit the S phase class as defined by the classifier set. This analysis was repeated with independent classifier sets for each S phase class. The total number of images

from each sample pool (auxin treated or control, 4, 7, or 10 hour timepoints) that fit the class in for each classifier set run was reported and plotted.

Computational analysis

**1. Repli-Seq read mapping and RT calculation**

Reads were aligned to the human genome (hg38) with STAR [26] (parameters: intron max - 1; align end type - EndToEnd; filter scores < 30; filter mismatches > 3; filter multimap scores range - 251; filter multimap n > 1). SamTools was used to filter uniquely-mapping reads that passed vendor QC checks and were not PCR duplicates (-F 3840; -q 30). These reads were binned into 1 kilobase windows and these windows were filtered to remove signal spikes (Z score > 2.5). Reads were then binned into 50 kilobase windows. A mappability score from Karimzadeh *et al* was used to mask unmappable regions of the genome [27]. Early over late replication timing (E/L RT) was calculated as the log2 (S phase bin 1/S phase bin 3) reads for the asynchronous sample and log2 (4 hour sample/10 hour sample) for the synchronized replicates. Quantile normalization was applied such that the distribution of replication timing values for two samples was matched. Loess smoothing with a span of 300 kilobases was applied to the normalized E/L RT values.

Weighted RT values were calculated as the weighted average of the four S phase bins for asynchronous values, or three S phase bins for the synchronized samples.

Analysis was carried out in R [28] and data were visualized using ggplot2 [29] and the UCSC Genome Browser [30].

## 2. Repli-Seq further analysis

Repli-Seq data for wildtype HCT116 cells (Accessions: 4DNFI5BZJXDE, 4DNFIC4VUF86) and IMR90 cells (Accessions: 4DNFILOYZWEM, 4DNFIKQDQCNB) were gathered from the 4D Nucleome data portal [31]. A Pearson correlation was used to quantify the degree of similarity between the profiles.

A Z score was calculated based on the difference between treated and control replication timings to identify significantly changed bins ($p < 0.05$). A false discovery rate (FDR) correction was applied to $p$ values to correct for multiple testing.

Publicly accessible data for histone marks, CTCF binding sites, and other genome characteristics were extracted from ENCODE [32] and the UCSC Table Browser [33] [Datasets Table]. ProSeq data from Rao *et al* [34] were used to identify genes with differential expression in response to cohesin loss upon auxin treatment. Briefly, significantly changed genes were classified as those with an absolute fold change greater than 2, an adjusted $p$ value less than 0.05, and at least 0.5 reads per kilobase million (RPKM). These thresholds yielded 74 genes that were induced upon auxin treatment and 16 that were downregulated. The genomic coordinates for these genes were extracted using BioMart [35]. BEDOPS was used to calculate the mean weighted RT for auxin-treated or control samples across annotated genomic ranges [36].

The 3D Genome Browser [37] was used to visualize and extract TAD calls from control and auxin-treated HCT116-SCC1-mAID cell HiC data from Rao *et al* [34]. This yielded a list of 2,307 contact domains in untreated cells and 883 domains in

auxin-treated cells. Juicer [38] was used to calculate compartment boundaries from untreated HCT116-SCC1-mAID cell HiC data. TAD boundaries that fell within 500 kilobases of a compartment boundary were considered separately to control for the effect of compartment status on replication timing. Compartment-independent TADs were classed as “SCC1-dependent” if they were not present (within 500 kilobases) in the contact domains of auxin-treated HCT116-SCC1-mAID HiC data, or “SCC1-invariant” if they remained in the contact domains of auxin-treated cells.

### 3. Spike-in ChIP-Seq read mapping and binning

Sequenced reads were aligned, filtered, and mapped using a custom written pipeline. Briefly, bowtie2 [39] was used to align reads to a genome reference that combined the experimental genome (*S. cerevisiae,* SacCer3) and the spike-in genome (*S. bayanus*). Duplicates were marked with Picard Tools [40]. Reads were filtered using Samtools [41] for primary alignments that passed QC and were not duplicates (-F 3840) with a minimal mapping quality of 30 (-q 30). Reads were binned into windows of desired size using bedtools [42].

### 4. ChIP-Seq normalization and calibration

The total number of reads mapping to the *S. cerevisiae* or *S. bayanus* genomes from each library was calculated. The reads mapping to *S. cerevisiae* were then calibrated by normalizing to the overall library depth (each bin was reported as the number of reads per 1 million mapped reads for the library). The sequencing depth normalized read counts were then calibrated to the spike-in following Hu *et al.* 2015 [43]:

$$\frac{R_{spike-in}^{input} * R_{expt}^{IP}}{R_{spike-in}^{IP} * R_{expt}^{input}}$$

where $R$ is the total number of reads mapping to either the spike-in $(R_{spike-in})$ or experimental $(R_{expt})$ genome for a given sample when both the input and IP aliquots for that sample were sequenced. When only the IP aliquot was sequenced, the calibration was modified to use the cell numbers in lieu of input read counts. The calibration became:

$$\frac{C_{spike-in} * R_{expt}^{IP}}{R_{spike-in}^{IP} * C_{expt}}$$

where $R$ is as before, and $C$ is the cell number for experimental and spike-in cultures that were mixed and carried forward to chromatin extraction.

## 5. Comparative datasets

In addition to the sequencing datasets generated as a part of this experimentation, previously generated data were used to validate and further explore the findings herein. Sources for additional datasets are as follows:

**Datasets Table.** Externally generated data used for the purpose of comparison, validation, or further analysis of the data generated in this work.

| Name | Dataset | Further Analysis | Citation/ Accession |
|---|---|---|---|
| BrdU-IP in HU | GSM3450327_130minHU_1000.bedGraph | NA | [44] |

| | | | |
|---|---|---|---|
| RT VGY129 | Sort-Seq data on VGY129 from A. Wallis (2018/10/06, VGY129, S phase and stationary) | RT calculation using RepliScope [45] | Personal communication |
| RT HCT116 | Processed Repli-Seq replication timing data from HCT116 cells | NA | [46] |
| RT HCT116 | Processed Repli-Seq replication timing data from HCT116 cells | NA | 4DNFI5BZJXDE, 4DNFIC4VUF86 [31] |
| RT IMR90 | Processed Repli-Seq replication timing data from IMR90 cells | NA | 4DNFILOYZWEM, 4DNFIKQDQCNB [31] |
| TAD calls HCT116 | HiC data from SCC1-mAID HCT116 cells with and without auxin-induced cohesin degradation | Pulled TAD and compartment calls with 3D Genome Browser [37] | [34] |
| SCC1 loss-induced // SCC1 loss-repressed genes | PRO-Seq data from SCC1-mAID HCT116 cells with and without auxin-induced cohesin degradation | Assessed differentially expressed genes | [34] |
| CpG islands | UCSC CpG Island track | NA | [47] |
| DNAse hot spots | DNase HS - DNase I Hypersensitivity in 95 cell types from ENCODE | NA | [48] |
| CTCF sites | ENCODE: CTCF ChIP-Seq in HCT116 | NA | ENCSR000BSE |
| Enhancers | Bed file of enhancer and super enhancer annotations in HCT116 cells | NA | [49] |
| EZH2 | ENCODE: EZH2 ChIP-Seq in HCT116 | NA | ENCSR046HGP |

| | | | |
|---|---|---|---|
| H4K20me1 | ENCODE: H4K20me1 ChIP-Seq in HCT116 | NA | ENCSR474DOV |
| H3K27ac | ENCODE: H3K27ac ChIP-Seq in HCT116 | NA | ENCSR000EUT_ENCFF083XMT |
| H3K36me3 | ENCODE: H3K36me3 ChIP-Seq in HCT116 | NA | ENCSR091QXP |
| H3K27me3 | ENCODE: H3K27me3 ChIP-Seq in HCT116 | NA | ENCSR810BDB |
| H3K4me1 | ENCODE: H3K4me1 ChIP-Seq in HCT116 | NA | ENCSR161MXP |
| H3K4me2 | ENCODE: H3K4me2 ChIP-Seq in HCT116 | NA | ENCSR794ULT |
| H3K4me3 | ENCODE: H3K4me3 ChIP-Seq in HCT116 | NA | ENCSR333OPW |
| H3K9ac | ENCODE: H3K9ac ChIP-Seq in HCT116 | NA | ENCSR093SHE |
| H3K9me2 | ENCODE: H3K9me2 ChIP-Seq in HCT116 | NA | ENCSR555LYM |
| H3K9me3 | ENCODE: H3K9me3 ChIP-Seq in HCT116 | NA | ENCSR179BUC |
| Super enhancers | Bed file of super enhancer annotations in HCT116 cells | NA | [49] |

## 6. Gene set analysis

The set of all annotated *S. cerevisiae* genes was retrieved from the *Saccharomyces* Genome Database (S288c assembly R64-1-1, gff version 3). Mitochondrial genes were removed from analysis. The mean normalized, spike-in calibrated value across the annotated gene region for each dataset was calculated in R and values plotted with ggplot as points and a boxplot for each gene in a given gene set or a boxplot for the full set of annotated genes.

**Gene sets**

gal: *GAL1*, *GAL2*, *GAL3*, *GAL4*, *GAL7*, *GAL10*, *GAL11*, *GAL80*

housekeeping: *ADH2*, *TAF10*, *TFC1*, *PDC1*, *PDA1*, *GPH1*, *ALG9*, *ACT1*

pol2: *POL2*

## 7. Beacon calculus simulation

Beacon calculus was used to simulate progression through S phase [50]. Origin parameters (location, efficiency, and mean replication timing) were taken from Hawkins, *et al* [51] and 500 simulations were run for each chromosome. All beacon calculus simulations and subsequent analyses were done in 1 kilobase windows.

### *Fork age*

The output of the beacon calculus simulations was read into R. For each bin across the genome, the time in S phase when the bin was synthesized was extracted. The difference between this time and the time of initiation of the replisome responsible for the synthesis was taken as the fork age. This was carried out for each of the 500 simulations, and the median age was reported as the fork age metric for that genomic bin.

### *Polymerase localization*

To calculate the expected polymerase locations at each stage of S phase, the replisome progression data from each of the 500 simulations for each chromosome were pooled. The replisome progression data was subset to pull out replisome positions that fell within a given time window of S phase (*i.e.* first half of S phase, first 10 minutes of S phase, etc). These data were plotted as a density of replisomes across the chromosome and a kernel density estimation fit to set a

probability of replisome presence at each genomic window for the given S phase fraction.

# Chapter III. Cohesin-mediated genome architecture does not define DNA replication timing domains

## Introduction

This chapter is based around work that was published as “Cohesin-mediated genome architecture does not define DNA replication timing domains” [52]. This ends with a discussion that includes perspective on the work published by other labs since the completion of work carried out herein.

### Genome replication and architecture

The double helix structure of DNA immediately lends itself to a mechanism for genome replication so simple and so elegant that the question of faithful transmission of hereditary information, to some, seemed solved as soon as it was posed [53]. However, behind the elegance of the underlying mechanism, the process of genome replication is in fact a complex, high-stakes, and intricately regulated cellular undertaking. A significant factor necessitating regulation is the sheer scale of the process: in a human cell, 12 billion bases are replicated in about 8 hours, with replication initiating from somewhere between 12,000 and 250,000 replication origins spread across 23 chromosome pairs [54]. Intriguingly, not all of these origins fire in any given S phase. Of those that do fire, not all fire at the same time. Instead, clusters of contiguous origins appear to fire in synchrony, leading to zones of hundreds of kilobases to megabases of DNA making up distinct replication timing domains (RD) [55]. The patterns of replication timing domains, driven by selective origin firings, lead to nuanced genome-wide replication timing profiles distinct to cell type and differentiation status [56–59] [Figure 2].

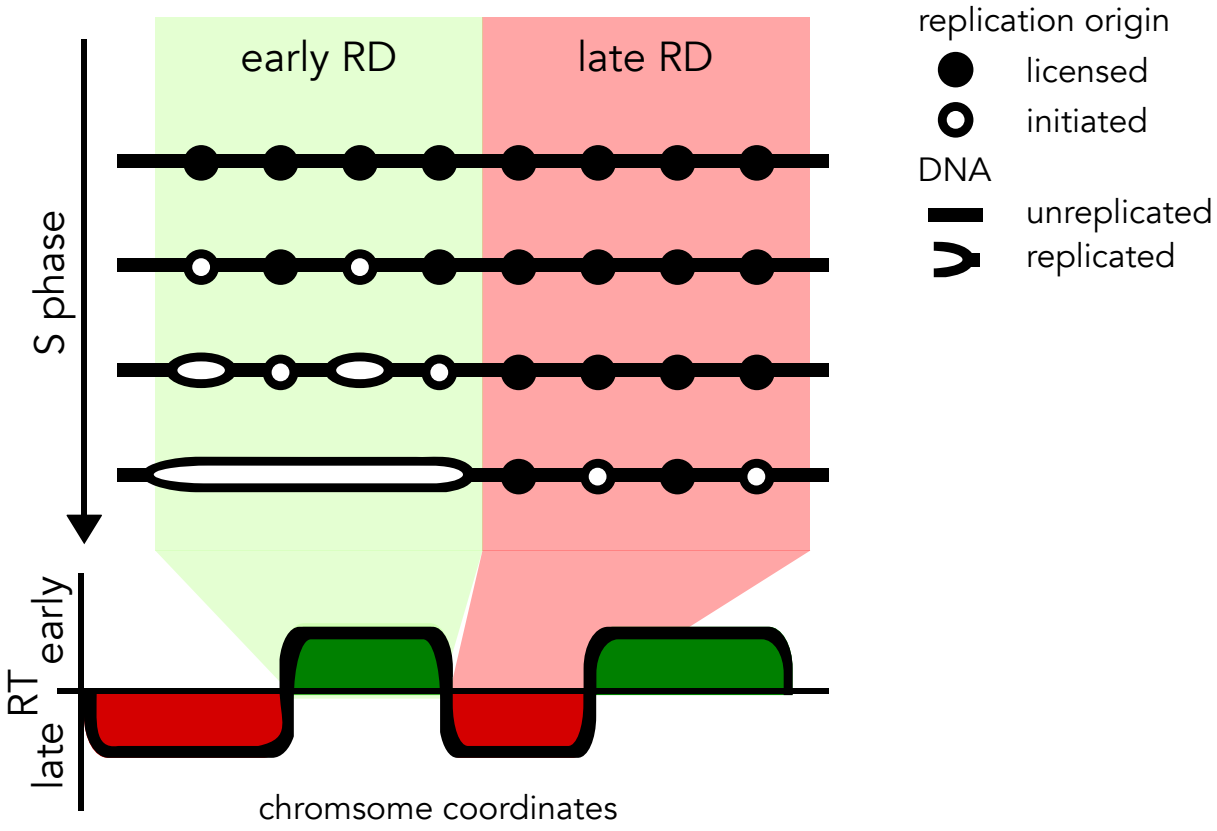



**Figure 2. Characteristic replication timing (RT) profiles emerge from selective firing of sets of origins at particular times in S phase.** More origins are licensed for replication than actually fire in any given S phase. Those that do fire tend to do so in clusters, resulting in zones of kilobases to megabases of DNA that are replicated at distinct times during S phase. These zones with characteristic replication timings are termed Replication Timing Domains (RD).

## Genome architecture as a determinant of replication timing

Work to characterize the cellular determinants of replication timing has uncovered a strong correlation to genome organization. This has been assayed through both cytological studies of replication foci over the course of S phase [60–64] and chromosome conformation capture studies [65–68]. Genome structure appears primarily to emerge from two fundamental organizing principles: (1) a loop-based organization dependent upon the proteins cohesin and CTCF results in domains of ≤1 Mb that are defined by enriched internal interactions and known as topologically associating domains (TADs) [69–71], and (2) compartmentalization into like chromatin states results in alternating 1–10 Mb domains of active or inactive

chromatin [69,72,73]. Of these different classes of genome structure, there is evidence that replication timing domains align well with compartments: early-replicating loci are significantly enriched for marks of open and active chromatin, or compartment A, whereas late-replicating loci share marks with compartment B [55,74] [Figure 3].

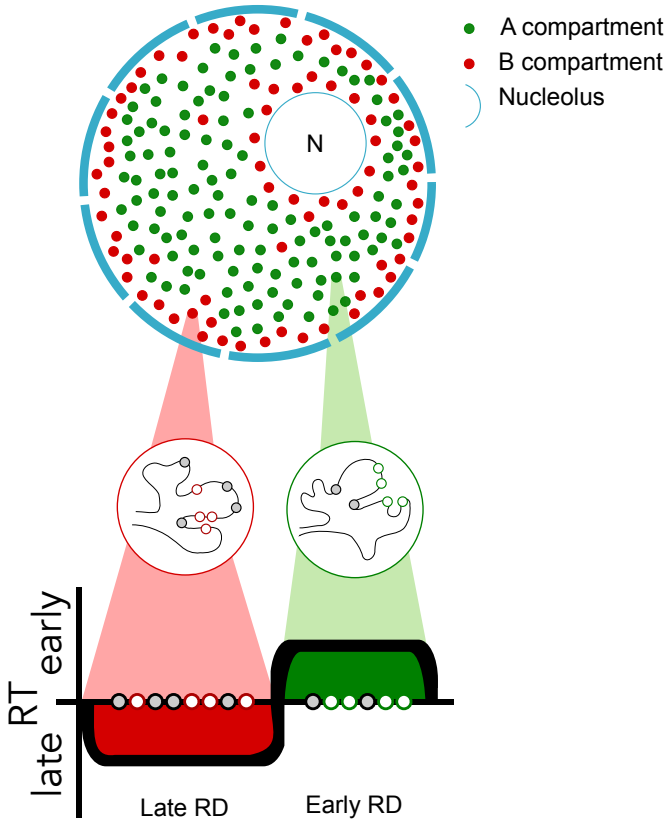



**Figure 3**. **Replication timing domains correlate with genome topology compartments**. Figure adapted from [75]. In mammals, early and late replication timing domains of around 1 megabase cover clusters of origins that are concordantly activated (green and red open circles) or passively replicated (grey filled circles). Early replication domains correlate with A compartments (euchromatic transcribed regions with active histone marks), whereas late replication domains correlate with B compartments (heterochromatin at the nucleolar and nuclear periphery with repressed histone marks).

Beyond the concordance between replication timing domains and genome compartments, the smaller TADs offer a compelling possibility for a unit of replication timing domain regulation. Their boundaries were found to map almost one-to-one to replication timing domain boundaries [58], and cohesin was found to

be enriched at origins as well as at TAD boundaries [7]. A link between TAD-based genome structure and replication origin function was further supported by the finding that both are established concomitantly in early G1 [76,77] and change correspondingly over the course of cellular differentiation [78]. The loop-based mechanism by which cohesin is hypothesized to drive genome organization is furthermore in agreement with work showing a correlation between chromatin loop size and interorigin distance: the relationship held as reprogramming of differentiated cells to embryonic cells resulted in both a shortening of interorigin distances and chromatin loop domains [79]. Together, this abundance of correlative evidence has made a convincing case for a "replication domain model" in which the stable and spatially distinct structural unit of TADs gives rise to the stable functional unit of replication timing domains [58,80].

Mechanistically, chromatin looping could drive replication timing by facilitating the passage of initiation factors which are present in the nucleus at levels sub-stoichiometric to the number of licensed origins. The limited abundance of these key activating factors has been shown in budding yeast to drive temporal replication timing patterns. Overexpression of sets of these factors – including CDK substrates Sld2 and Sld3, along with binding partner Dpb11, Cdc45, and the kinase DDK – allows for early firing of characteristically late-firing origins [13,81]. The phosphorylation of MCM by DDK has emerged as a particularly rate-limiting step for the activation of licensed origins [75]. Factors dictating the proclivity of certain origins to access limiting initiation factors over others at particular times in S phase have been elucidated to some extent across various model organisms. Origins made less accessible by closed chromatin fire later [74] and the kinetochore protein Ctf19 preferentially recruits DDK to pericentromeric origins in *S. cerevisiae*

[82]. In budding yeast, transcription factors Forkhead 1 and Forkhead 2 bind certain origins and recruit DDK, promoting early firing of Forkhead-bound origins [21,83]. Across yeast and mammals, Rif1 works to suppress early firing of origins, including those at yeast telomeres, by recruiting protein phosphatase 1 (PP1), which counteracts DDK phosphorylation [84–88]. However, these specific *trans* interactions cover a relatively small fraction of the genome: for much of the genome, the composite set of known factors still does not account for all of the observed nuances of replication timing profiles. Thus, the hypothesis that spatial clustering of licensed origins allows for efficient recycling of limiting factors from one origin to the next – facilitating temporally clustered firing of origins spatially clustered in replication factories – offers a mechanism by which genome architecture could drive replication timing.

## Implications of replication timing

The conservation of replication order suggests an importance to cell viability - this is borne out in the array of consequences resultant from misregulated replication timing. Late-replicating regions of the genome have been shown to have higher mutation rates [89,90]. Chromosomal aberrations are also associated with replication timing: somatic copy number alteration (SCNA) boundaries were found to be enriched in late-replicating regions over early in human cells [91]. Interestingly, while late-replicating regions showed more copy number variation boundaries overall, late-replicating regions were particularly associated with deletions, whereas amplifications were found to be slightly more likely in early-replicating regions [91]. While the replication timing of a particular locus may impact the chances for mutations or copy number changes in that locus, loss of replication timing control genome-wide can lead to overall increases in chromosomal aberrations. This is

suspected to be driven by replication stress: over-firing of origins was found to deplete nucleotide pools, driving replication checkpoint activation [81]. It is suspected that this replication stress is a driver of chromosomal instability in early cancer stages [92], and intriguingly, cell supplementation with nucleosides was shown to ameliorate chromosome segregation errors in cancer cell lines [93] and to decrease DNA damage and oncogene-induced transformation events [92].

Outside of primary genome sequence, replication timing is also suspected to play a role in epigenome maintenance. Notably, replication timing changes in cancer cells were found to correlate to changes in epigenetic marks and DNA methylation [94]. While the direction of the relationship between epigenetic state and replication timing can be challenging to parse, a role for replication timing as a driver of transcriptional state has been shown in an experiment employing nuclear injection of DNA during S phase. Irrespective of the sequence delivered, DNA injected in early S phase was transcribed to a higher level than DNA injected late in S phase [95]. A similar microinjection approach was used to show that DNA introduced early in S phase was more likely to be packaged with acetylated histones, conductive to active transcription, than the same DNA introduced in late S phase [96].

Finally, replication timing can be a powerful method for tuning the expression level of proteins required in high abundance during the process of DNA replication of new-genome packaging. While a replication-dosage compensation pathway exists to buffer cells from the transient copy number changes that necessarily occur during ongoing DNA replication of the genome from one to two *n* in an S phase cell [97], in *S. cerevisiae,* histone genes H2A and H2B escape this control to show increased expression following their replication early in S phase [98]. Changing

replication timing of the H2A and H2B histone genes from early to late resulted in diminished mRNA abundance over S phase [98]. This finding in yeast describes an elegant cellular mechanism to meet the high S phase-specific needs for replication-associated products such as histones by exploiting the copy number increase inherent to the process of genome replication. An analogous mechanism may be at play in mammalian cells, adding to the list of potential consequences of unregulated replication timing.

## Cohesin

Cohesin is a multi-protein complex essential for accurate chromosome segregation [99]. The complex was found to exist in a large ring conformation and it was shown that this ring encircles DNA duplexes, thus driving cohesion of sister chromatids after replication [100]. At the onset of mitosis cohesin-joined sister chromatids align at the cell equator, pulled to one or the other of the cell poles by their kinetochore's attachment to the microtubule organizing center. The cohesin entrapment of the two sister chromosomes allows for tension that acts as a signal for the metaphase checkpoint to assess that all chromosomes are duplicated and attached before proceeding through the next steps of mitosis. When all sister chromatids are aligned and the metaphase checkpoint satisfied, the cohesin rings are cleaved by separase, which releases chromatids to segregate to opposing cellular poles such that each daughter cell receives one full copy of the genome.

The role of cohesin in genome biology expands beyond facilitating accurate chromosome segregation. In *S. cerevisiae*, cohesin mutants were found to be defective in DNA damage repair – leading to the hypothesis that efficient repair of double strand breaks by homologous recombination is enabled by the close

proximity of a repair template due to sister chromatid cohesion [101]. Further work in mammalian cells has found that cells expressing a non-phosphorylatable mutant of cohesin subunit Smc1 show reduced repair efficiency and increased DNA damage sensitivity [102,103].

The cohesin ring has increasingly been implicated in regulating genome biology in interphase cells. This was first suspected through the observation that cohesin expression persisted across all stages of the cell cycle, and extended even to non-cycling cells such as post-mitotic neurons [99]. Further investigations suggested that cohesin could drive enhancer-promoter interactions, as studies in *Drosophila* showed cohesin loading to be required for the long-range activation of homeobox protein genes [104], for *Runx* expression in zebrafish [105], and for genes controlling embryonic stem cell identity in murine embryonic stem cells [106]. The cohesin complex was also implicated in gene silencing in budding yeast, as cohesin mutants showed loss of the silencing boundary, allowing SIR spreading from the silenced mating locus [107].

Further studies found the patterns of cohesin occupancy on the interphase genome to be distinct from the occupancy seen in chromatid cohesion; these distinct binding patterns shed light on cohesin's role in gene silencing. Cohesin binding sites were correlated heavily with genomic loci bound by insulator CCCTC-binding factor (CTCF) [108]. While not required for cohesin binding onto the genome, CTCF was required for cohesin enrichment at particular loci, and for the insulation of promoters from distant enhancers [108]. The advent of 'chromosome conformation capture' methods opened up the opportunity for high resolution analysis of long-range genome locus interactions and genome topology. These

methods uncovered the two orders of chromosome topology previously introduced, compartments and topologically associating domains (TADs) [109], and furthermore showed the CTCF and cohesin-bound sites to be enriched at strong topological domain boundaries [110]. This led to a model in which cohesin and CTCF drive the organization of the genome into distinct TADs. Given the concordance of replication domains with TADs and the observation of synchronously-firing clusters of origins within replication domains, cohesin emerged as a strong potential contender for driving replication timing, as cohesin-based looping could function to bring origins into spatial proximity, driving synchronous firing by facilitating efficient recycling of limiting initiation factors.

Auxin degron tag

Due to the essential role of cohesin in chromosome segregation, studies investigating the causal role of cohesin on genome architecture – and its role in replication timing – were initially limited by an inability to carry out a clean cohesin knock-out or complete depletion over the course of a single cell cycle. A study which used cohesin knockdown by small interfering RNA (siRNA) showed a reduction in active origins and an increase in interorigin distance in response to depleted cohesin levels [7]. However, this work may have been limited by the slow approach of regulating cohesin at the transcriptional level, and furthermore did not explore an effect on replication timing domains. The development of the auxin-inducible degron system [111,112] has made this sort of research much more tractable, allowing for a rapid and near-complete depletion of cohesin [Figure 4].

The auxin-inducible degradation system exploits the fact that while SCF ubiquitin ligases are universal across eukaryotes, their target specificity is driven by F box

proteins which can be specific to certain phyla [111]. TIR1 is a plant-specific F box protein which forms an E3 ubiquitin ligase induced by the phytohormone auxin. Thus, by introducing TIR1 to a yeast or mammalian cell and adding a TIR1 recognition tag (auxin-inducible degron, AID tag) to a gene of interest in that cell line, it is possible to engineer specific and rapidly inducible degradation of that gene product.

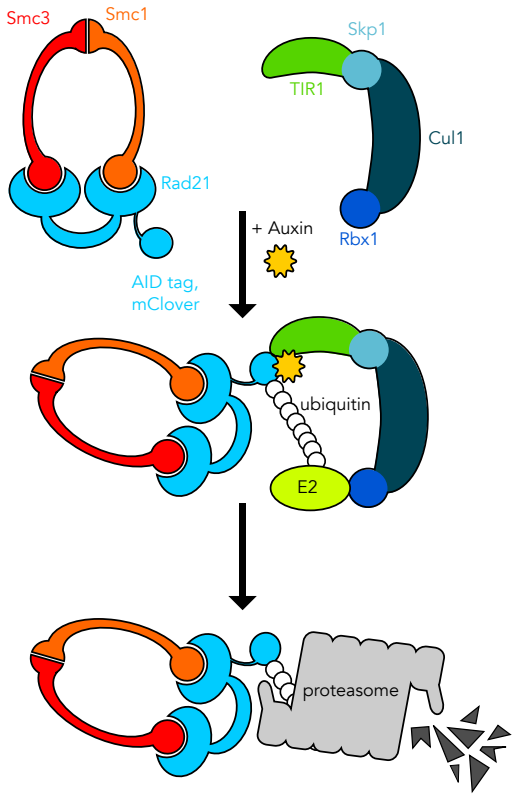



**Figure 4**. **Auxin inducible degradation with plant ubiquitin F box protein TIR1**. Figure adapted from [112].

Taking advantage of the auxin inducible degron system, chromosome conformation studies have been carried out following acute cohesin ablation via a degron tag on cohesin's SCC1 subunit [34]. Strikingly, these studies showed a rapid loss of TADs upon cohesin degradation, while compartment structure remained intact – and even strengthened, as loop domains that span multiple compartments

were no longer actively driving association of genetic loci of different states [Figure 5] [34]. Given previous findings of cohesin's role in transcription regulation and epigenetic boundary demarcation, cohesin depletion and the genome-wide loss of TADs was anticipated to result in widespread changes in gene expression level and epigenetic states. Surprisingly, nascent transcription, as analyzed by precision nuclear run-on sequencing, showed only modest changes: 216 genes (1% of all unexpressed in the control condition) underwent ectopic activation in the absence of cohesin and 64 genes (0.5% of all expressed in the control condition) showed a 2-fold change in expression level in the absence of cohesin [34]. Similarly, epigenetic states did not demonstrate the anticipated changes, as the ChIP-Seq profiles for a wide range of histone marks (H3K4Me3, H3K4me1, H3K27Ac, H3K27me3, H3K9me3, H2.AZ, H3K36me3, H3K16Ac, H3K79me2) were not significantly different in the auxin-treated versus control samples.

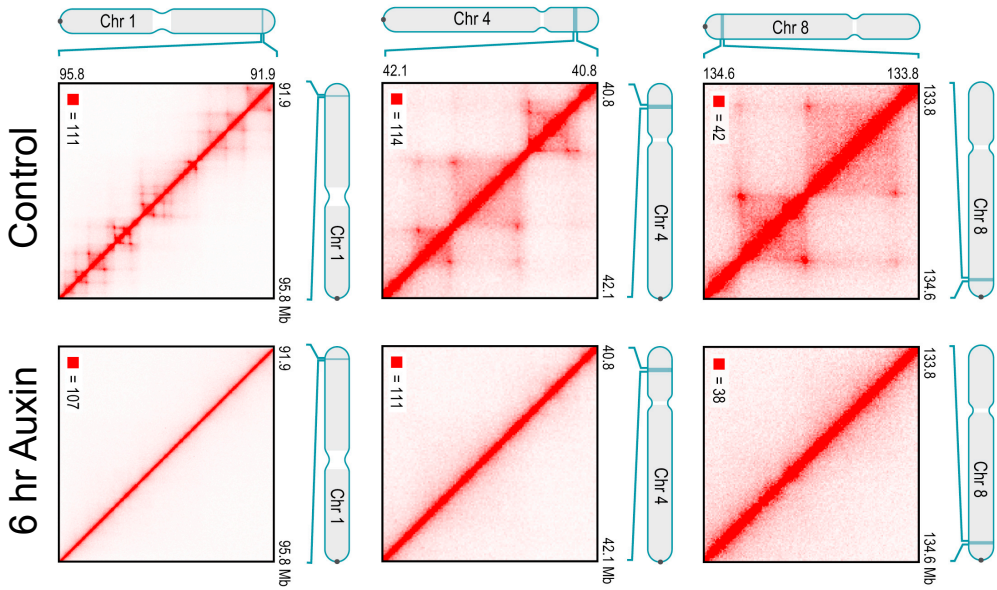



**Figure 5**. **Auxin-induced cohesin depletion resulted in a rapid loss of TAD architecture**. HiC contact matrices from control and cohesin-depleted (6 hours of auxin treatment) HCT116 SCC1-mAID cells, data from [34]. Three representative loci are shown, with darker red representing more abundant contacts. TADs - the squares of enriched contacts –are lost upon auxin treatment. Figure adapted from [34].

Motivation

In this study, an auxin-inducible degron tag on cohesin subunit SCC1 was exploited to allow for rapid degradation of cohesin. This system was used along with analysis of genome replication by sequencing and microscopy techniques to test whether the observation of a strong correlation between cohesin-mediated genome architecture and replication timing is substantiated as a causative relationship.

## Results

### System development and validation

#### *Efficient and rapid degradation of endogenous SCC1*

It has been previously shown that degradation of the SCC1 subunit of the cohesin complex leads to an inability of the complex to associate with DNA [34]. Here, an HCT116 (human colon cancer) cell line was used in which the endogenous SCC1 was tagged with a mini-auxin inducible degron (mini-AID) tag to allow for rapidly inducible degradation of the cohesin subunit, and thus destabilization of the cohesin complex [112]. The SCC1 protein was further tagged with an mClover fluorescent tag to facilitate monitoring of the degradation efficiency. To confirm loss of the SCC1 fusion protein after 2 hours of auxin treatment, mClover signal was assessed in treated and control cells by microscopy [Figure 6A] and flow cytometry [Figure 6B]. These two methods yielded comparable values: after 2 hours of auxin treatment 73% of cells were mClover-negative by flow cytometry (n=40,649), and 78% were mClover-negative by microscopy (n=54) [Figure 6B,C]. Cells that retained residual mClover signal after auxin treatment had notably reduced intensities of signal relative to untreated levels [Figure 6A]. To further confirm that the loss of mClover related to loss of SCC1, immunofluorescence (IF) was carried out for SCC1. 81% of cells imaged were negative for SCC1 stain after 2 hours of auxin treatment and those cells that retained any SCC1 signal after treatment were markedly reduced in signal intensity [Figure 6A, C]. These results agree with previous work that has demonstrated a loss of SCC1 within 1 hour of auxin treatment in the same HCT116 cell line [34,112].

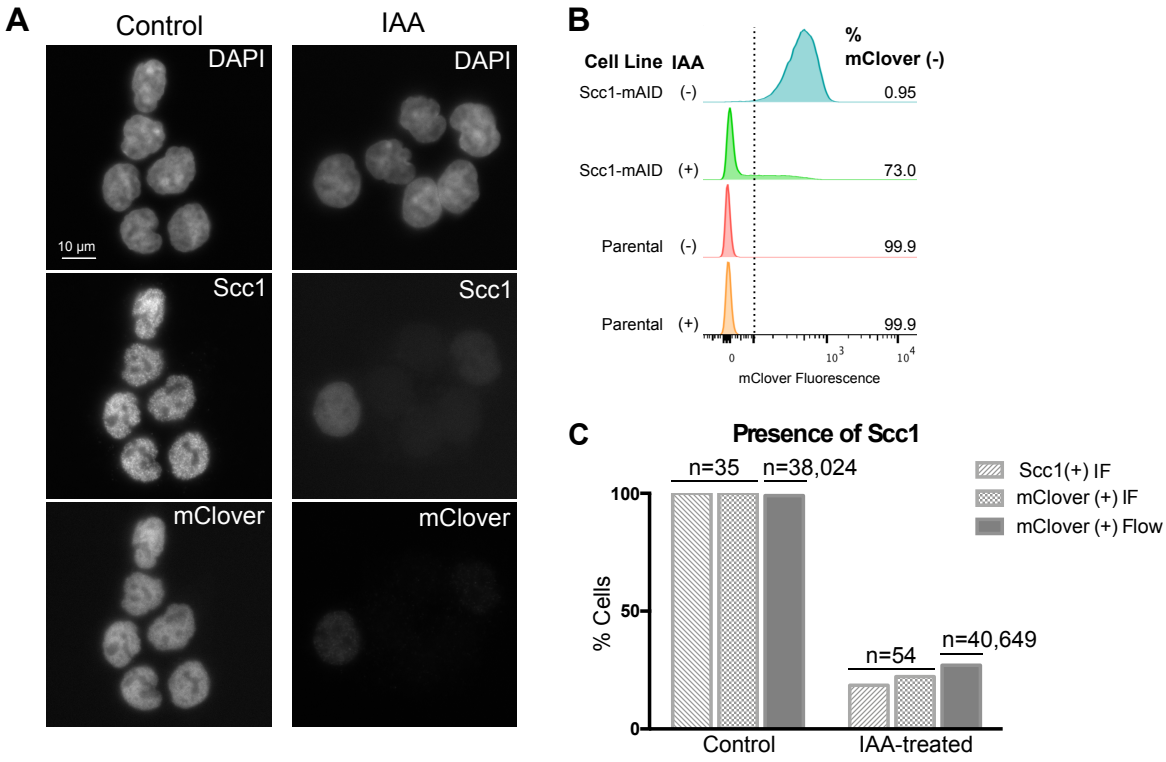



**Figure 6**. **Auxin treatment results in loss of SCC1**. **(A)** Immunofluorescence for SCC1 and mClover in HCT116 SCC1-mAID-mClover cells under control conditions or 2 hours of auxin treatment. **(B)** Flow cytometry analysis for mClover fluorescence using both HCT116 SCC1-mAID-mClover cell line and parental HCT116 cell line as a negative control. The mClover-positive gate was set based on the parental cell line at $10^2$ fluorescence units. **(C)** Comparison of direct IF read-out of loss of SCC1 and indirect read out of loss of mClover through both flow cytometry and IF.

### *S phase progression in the absence of cohesin*

The cohesin ring has been implicated in various functions outside of its proposed role in genome topology through the organization of TADs. To ensure that any changes in replication timing seen were due to changes in genome topology, then, it was important to confirm that auxin-induced loss of cohesin over S phase did not change overall replication progression. This was assessed by synchronizing cells using a thymidine block into a nocodazole shake-off, and depleting cohesin or leaving cells as controls while monitoring progression through the subsequent S phase. Although bulk DNA content had not increased by 4 hours post release, the presence of BrdU-positive cells indicated that both cohesin-depleted and control cells had initiated DNA replication [Figure 7]. In all

subsequent monitored time points, the DNA content of BrdU-positive cells was closely matched between treated and control samples, with the exception of the 24 hour timepoint. While the control cells evidently continued to cycle (as indicated by cells of even higher BrdU content) the auxin-treated sample no longer included BrdU-positive cells at 24 hours, likely because cycling cells did not survive mitosis in the absence of cohesin.

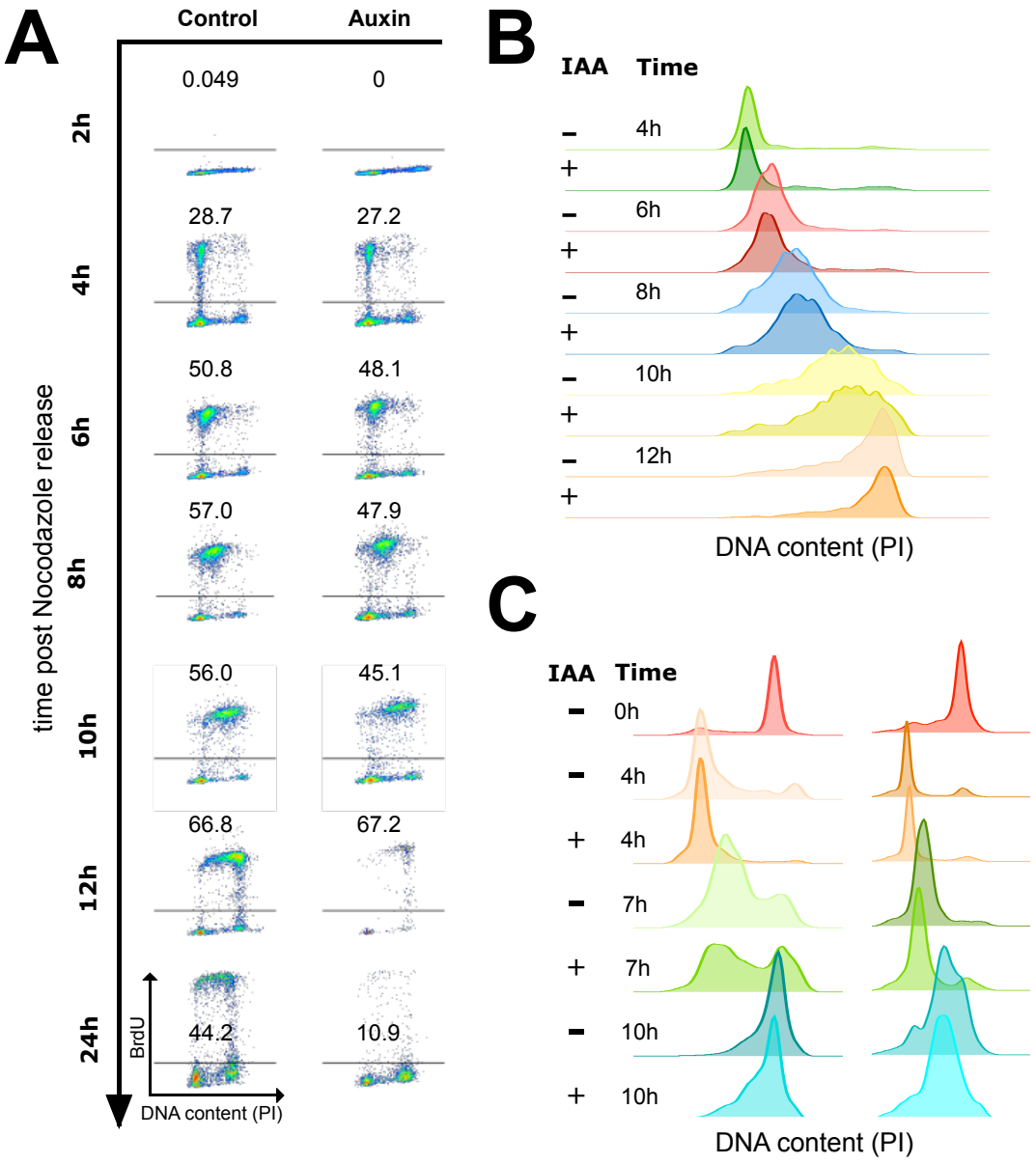



**Figure 7**. **Cohesin depletion does not dramatically alter overall S phase progression dynamics**. Cells were synchronized with a thymidine arrest followed by nocodazole and mitotic shake-off. Auxin was added upon nocodazole release and cells were pulsed with BrdU for 2 hours prior to harvest at 4, 6, 8, 10, 12, and 24 hours following nocodazole release. Cells were analyzed

with flow cytometry for DNA content (PI) and BrdU incorporation. **(A)** Scatterplot of DNA content (PI) versus BrdU signal for cells in each sample. The line shows the gate set for calling BrdU-positive versus BrdU-negative cells. The number shows the percent of cells that were BrdU-positive out of the singlet population. **(B)** Density plot showing the DNA content (PI) of the BrdU-positive cell population for each sample. **(C)** Two additional replicates of thymidine-nocodazole synchronization experiments with samples released into auxin or left as controls. BrdU was added at 2, 5, and 8 hours, followed by cell harvest 2 hours later (*i.e.* at 4, 7, and 10 hours). Cells were analyzed by flow cytometry, and the DNA content (PI) of Brdu-positive singlet cells is show as a density plot.

### *Establishing the Repli-Seq protocol*

Replication sequencing or 'Repli-Seq' is a widely used method for determining patterns of replication timing (RT) across the genome [21,25]. Briefly, in this method, asynchronously growing cells are pulsed with the DNA analogue BrdU, which can be incorporated in place of thymidine by replicative DNA polymerases [Figure 8]. Cells at particular stages of S phase (*i.e.* cells *early* in S phase which have just started DNA replication, or *late*, completing replication) can be collected with fluorescence activated cell sorting (FACS) with gates based on DNA content. Nascent DNA from each S phase fraction is then immunoprecipitated by pulling down BrdU. Libraries are prepared from the immunoprecipitated DNA and subjected to high-throughput sequencing. Replication timing across the genome is commonly calculated as the log score of the proportion of mapped sequencing reads from the early S phase cells over those from the late S phase cells for every genomic bin.

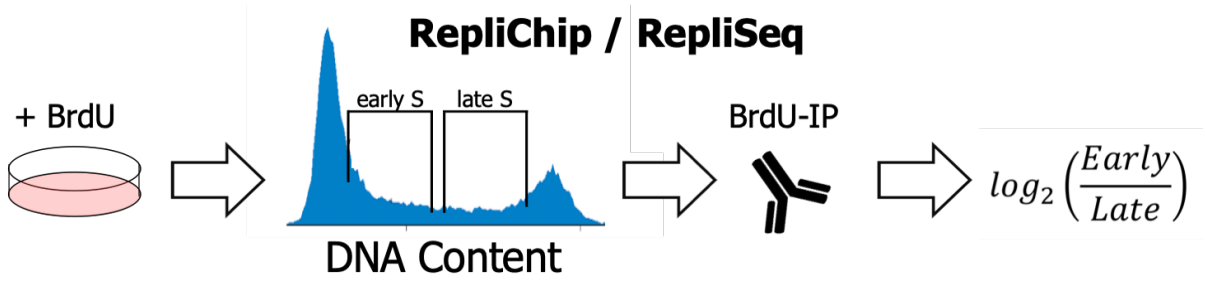



**Figure 8**. **Repli-Seq protocol schematic**. Asynchronously growing cells are pulsed with DNA analogue BrdU, which can be incorporated in place of thymidine in nascent DNA. Cells are processed by FACS on the basis of DNA content to select cells from early and late S phase. DNA

is extracted from the early and late S phase cell populations, and an immunoprecipitation is used to pull down nascent DNA which was synthesized during the BrdU pulse. This DNA can be analyzed by sequencing (Repli-Seq) or microarray (RepliChip). Replication timing is frequently calculated as the log of normalized read counts from the early S phase library over the normalized read count from the late S phase library.

In the course of this work, I established this method in the laboratory, with the modification of collecting four S phase bins by FACS to open the possibility of higher-resolution replication timing calculations across S phase [Figure 9]. The normalized read counts showed the progression of replication, with peaks apparent in the earliest S phase fraction (S1) widened in reads mapped from later S phase fractions (S2, S3) [Figure 9B].

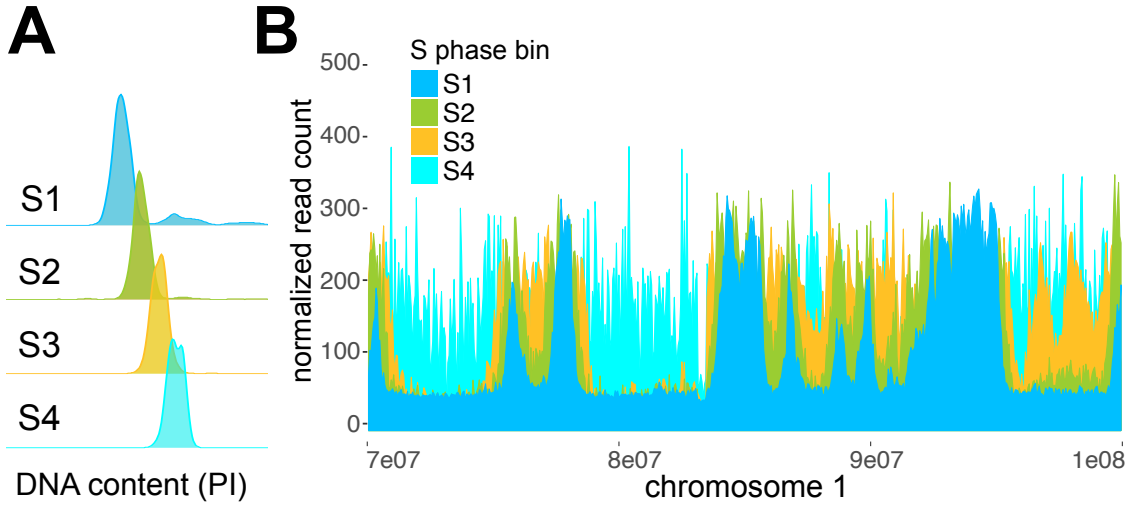



**Figure 9**. **(A)** FACS was used to select cells from four S phase bins, S1, S2, S3, and S4 progressing from early to late S. The FACS sort was confirmed with flow cytometry. **(B)** Normalized read counts from four S phase bins collected by FACS on asynchronous BrdU-pulsed cell culture. The read profiles show replication proceeding outward from peaks in the earliest S phase sample (S1).

To confirm that the protocol was yielding expected results, RT profiles for the control sample were calculated by taking the reads from the S1 fraction over those of the S4 fraction. This yielded a metric comparable to the RT profiles reported on the replication timing database, ReplicationDomain [46]. Thus, the results for this experiment were directly compared to previously generated RT profiles for HCT116 cells to validate that the method was generating expected

results [Figure 10]. The RT profile generated through this work showed distinct domains of early or late replication timing across the genome which aligned well with those seen in the ReplicationDomain dataset [Figure 10], and the two datasets showed a strong statistical correlation (Pearson's test, 0.903).

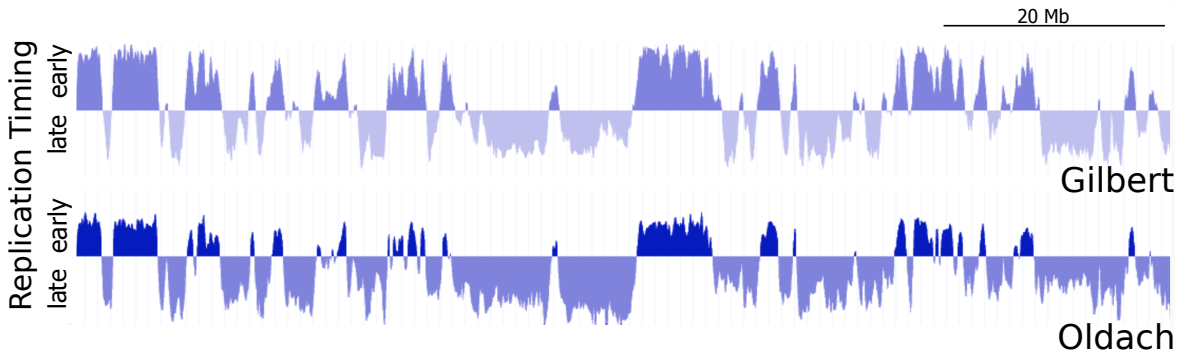



**Figure 10**. Comparison of Repli-Seq data generated in this experiment (Oldach, bottom) to previously-generated HCT116 Repli-Seq data from ReplicationDomain (Gilbert, top) [46]. Chr1:800,000-100,000,000.

## Loss of SCC1 does not perturb replication timing patterns

### *Asynchronous depletion of SCC1*

After establishing and validating the Repli-Seq protocol, this technique was used to assess the impact of cohesin loss on DNA replication timing. Asynchronously growing cells were treated with auxin for 2 hours or left as untreated controls, pulsed with BrdU for a further 2 hours, and then FACS enriched on DNA content into four S phase fractions and processed for Repli-Seq.

Surprisingly, the overall profile of replication timing did not change in response to auxin-induced SCC1-ablation [Figure 11A]. Quantitatively, the correlation between the control and treated profiles was comparable to the correlation between the control and the previously-reported HCT116 replication timing profile from ReplicationDomain [Figure 11C]. To facilitate identification of regions with changed replication timing, the difference of replication timing in control versus

auxin-treated samples was calculated for each genomic locus. Replication timing was only significantly changed (FDR-adjusted p < 0.05) in 0.34% of mappable genomic bins (189 bins out of 56,034 bins with mappability score > 0.5) [27]. Significant differences were enriched in bins with low mappability and dispersed across the genome (median distance between significant bins: 925,000 bases), suggesting that these differences were a consequence of experimental noise rather than domains of biological relevance [Figure 11B, Figure 13]. Furthermore, loci with significantly-changed replication timing were not enriched at TAD boundaries, as might be expected if these changes resulted from loss of TAD architecture [Figure 11D].

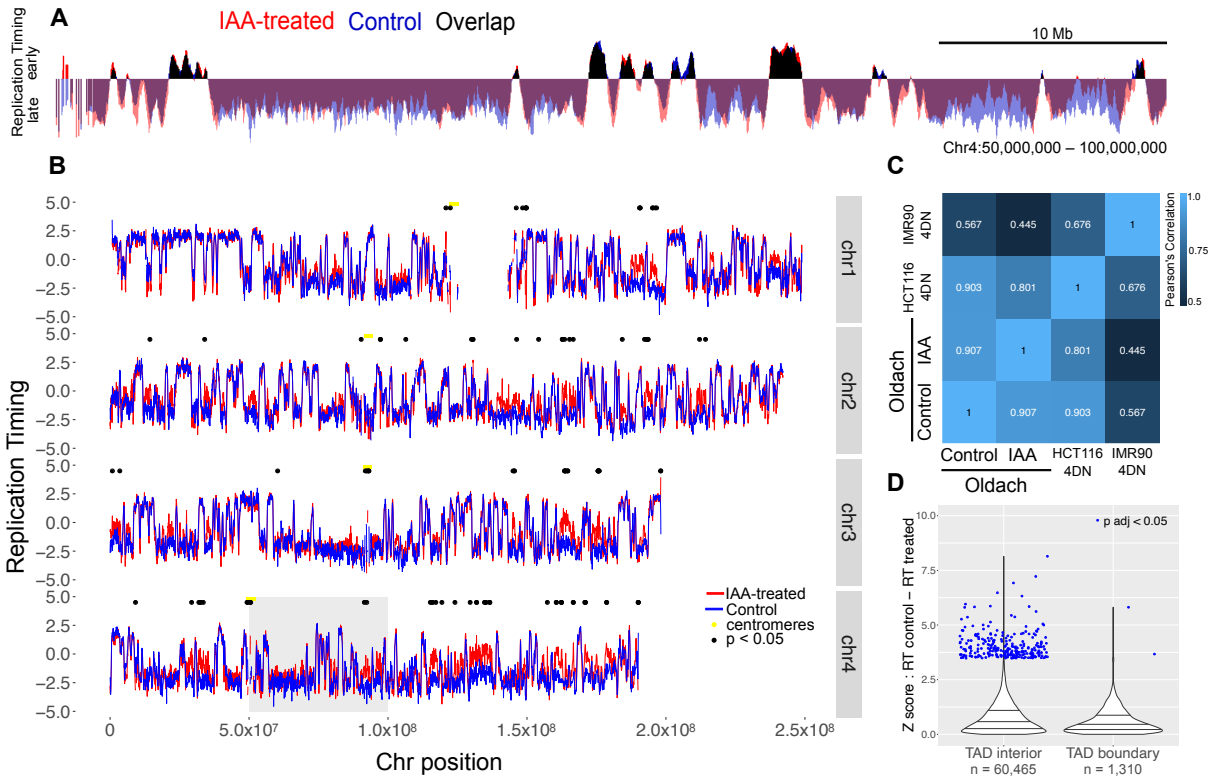



**Figure 11**. **Acute loss of cohesin in S phase does not perturb replication timing.** Asynchronously cycling cells were treated with 500 μM IAA for 2 hours, then pulsed with 100 μM BrdU for 2 hours. Cells in four S phase bins were collected by FACS, and DNA was extracted for Repli-Seq. **(A)** Example locus overlay of replication timing (RT, log-scaled ratio of read counts in early over late S phase bin) for auxin-treated versus control cells. **(B)** RT plots for four example chromosomes with significantly changed loci marked with black dots. The gray box highlights the region shown in (A). **(C)** Pearson correlation calculated for RT values across genome compared to previously published 4D Nucleome (4DN) Repli-Seq data. HCT116 RT data from the 4DN were

included with the expectation of similar RT traces, and IMR90 data were included as an outgroup. **(D)** Violin plot showing the distribution of Z scores for control versus auxin-treated RT difference across the genome, separating out genomic bins covering a TAD boundary.

As genome-wide analysis did not demonstrate widespread perturbations to the replication timing profile in response to the loss of cohesin, genomic regions marked by features hypothesized to show a change were specifically assessed. These included genomic regions with features characteristic of early replication (super enhancers, DNase hotspots, CpG islands, H3K4 methylated regions, H3K9 and H3K27 acetylated regions), late replication (H3K9me3, polycomb binding sites), CTCF binding sites, or genes with expression changes in response to auxin-induced loss of SCC1. None of these sets of sites showed a differential response [Figure 12].

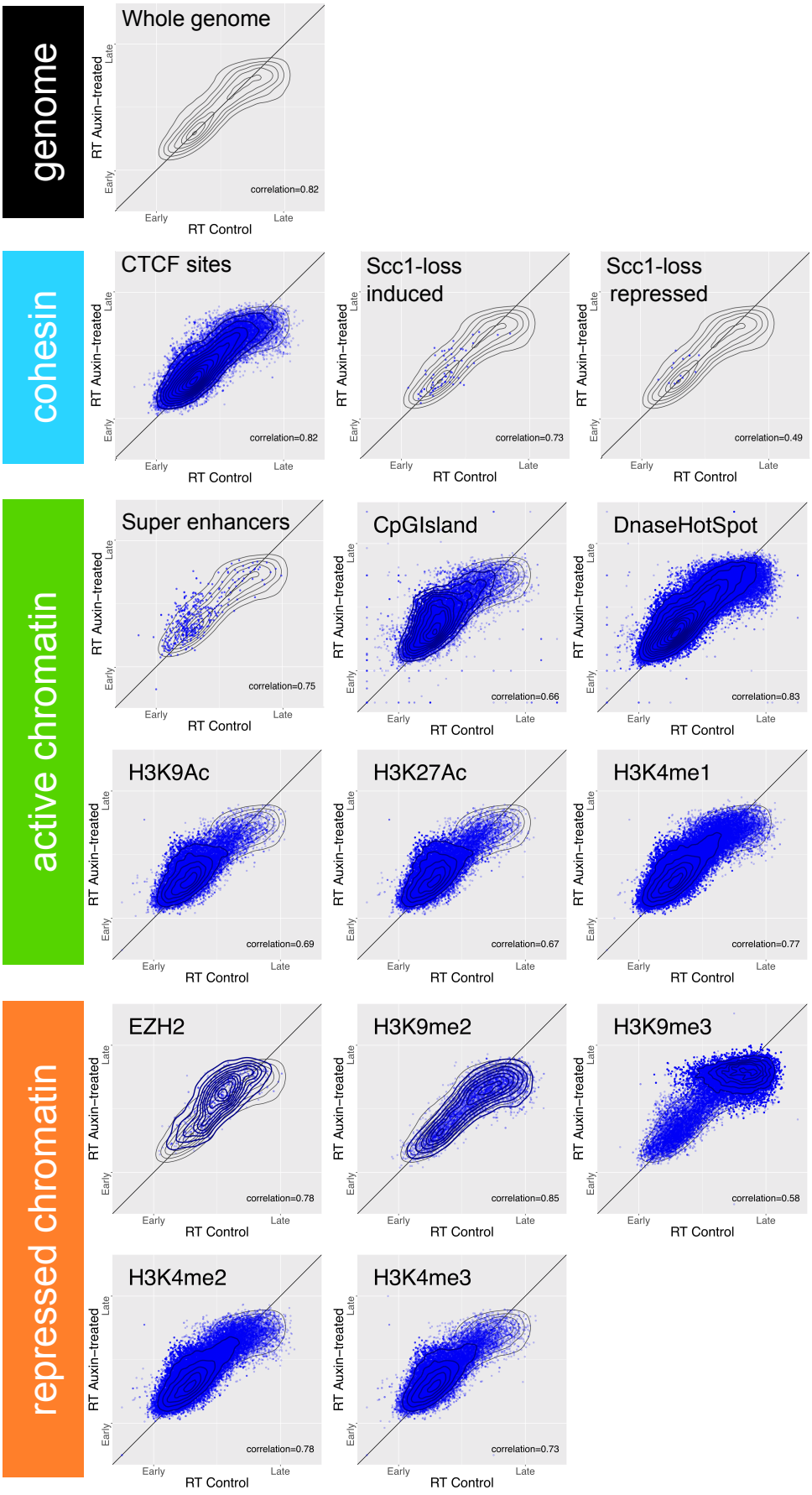

genome
Whole genome
correlation=0.82
cohesin
CTCF sites
correlation=0.82
Scc1-loss induced
correlation=0.73
Scc1-loss repressed
correlation=0.49
active chromatin
Super enhancers
correlation=0.75
CpGIsland
correlation=0.66
DnaseHotSpot
correlation=0.83
H3K9Ac
correlation=0.69
H3K27Ac
correlation=0.67
H3K4me1
correlation=0.77
repressed chromatin
EZH2
correlation=0.78
H3K9me2
correlation=0.85
H3K9me3
correlation=0.58
H3K4me2
correlation=0.78
H3K4me3
correlation=0.73
RT Auxin−treated
RT Control
Early
Late

**Figure 12. Correlation between control and auxin-treated asynchronous RT in genome regions associated with cohesin or replication timing.** For each bin in the genome, RT can be compared between the control and treated samples. Genome regions covering CTCF binding sites and the set of genes upregulated in response to cohesin-ablation (SCC1-loss induced) or downregulated in response to cohesin-ablation (SCC1-loss repressed) [34] were specifically assessed for changes in RT in response to cohesin ablation. Genome regions with histone marks and DNA binding sites associated with early replication timing (active chromatin) or late replication timing (repressed chromatin) were also assessed. For each of these genome features, the external datasets used to define genome regions are listed in Materials and Methods. The whole-genome behavior is plotted as a grey topology for reference. The dark blue points and topology indicate the RT for genomic bins annotated for each characteristic or genome feature.

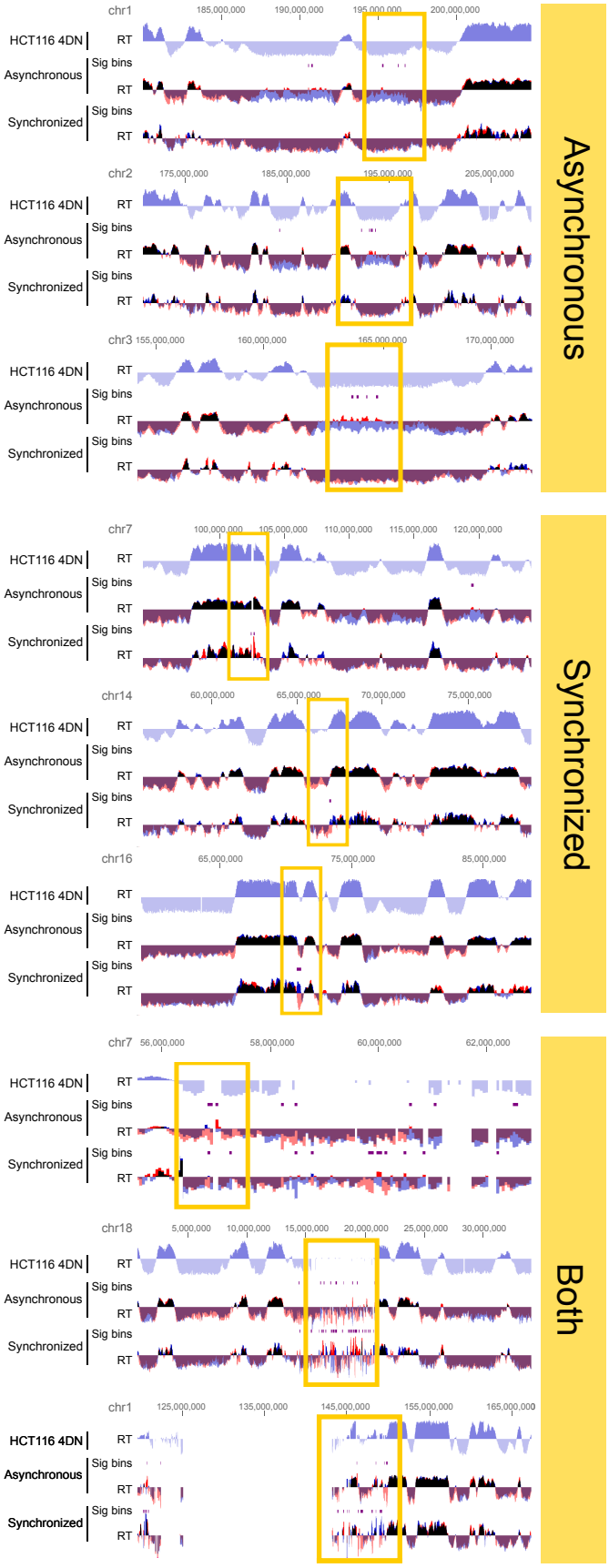



**Figure 13**. **Example loci showing significant RT differences**. Loci were annotated for significant differences ($p < 0.05$ after FDR correction for multiple testing) between replication timing in auxin-treated versus control samples under experimental conditions of cohesin ablation in

asynchronous cells (top), cohesin ablation from G1 in synchronized cells (middle), or both (bottom). Publicly available HCT116 replication timing data (4DN) shown as a reference.

To investigate the impact of cohesin ablation on the cytological patterns of replication timing, auxin-treated or control BrdU-pulsed SCC1-mAID-mClover cells were classed into five S phase categories based on BrdU foci patterning [Figure 14A] [61]. The treatment with auxin and resultant loss of cohesin did not alter the abundance of cells across each of the five classes [Figure 14B]. In combination with the finding that auxin-induced loss of SCC1 does not change progression through S phase, this suggests that auxin treatment does not change the spatial patterns of replication characteristic to different stages of S phase.

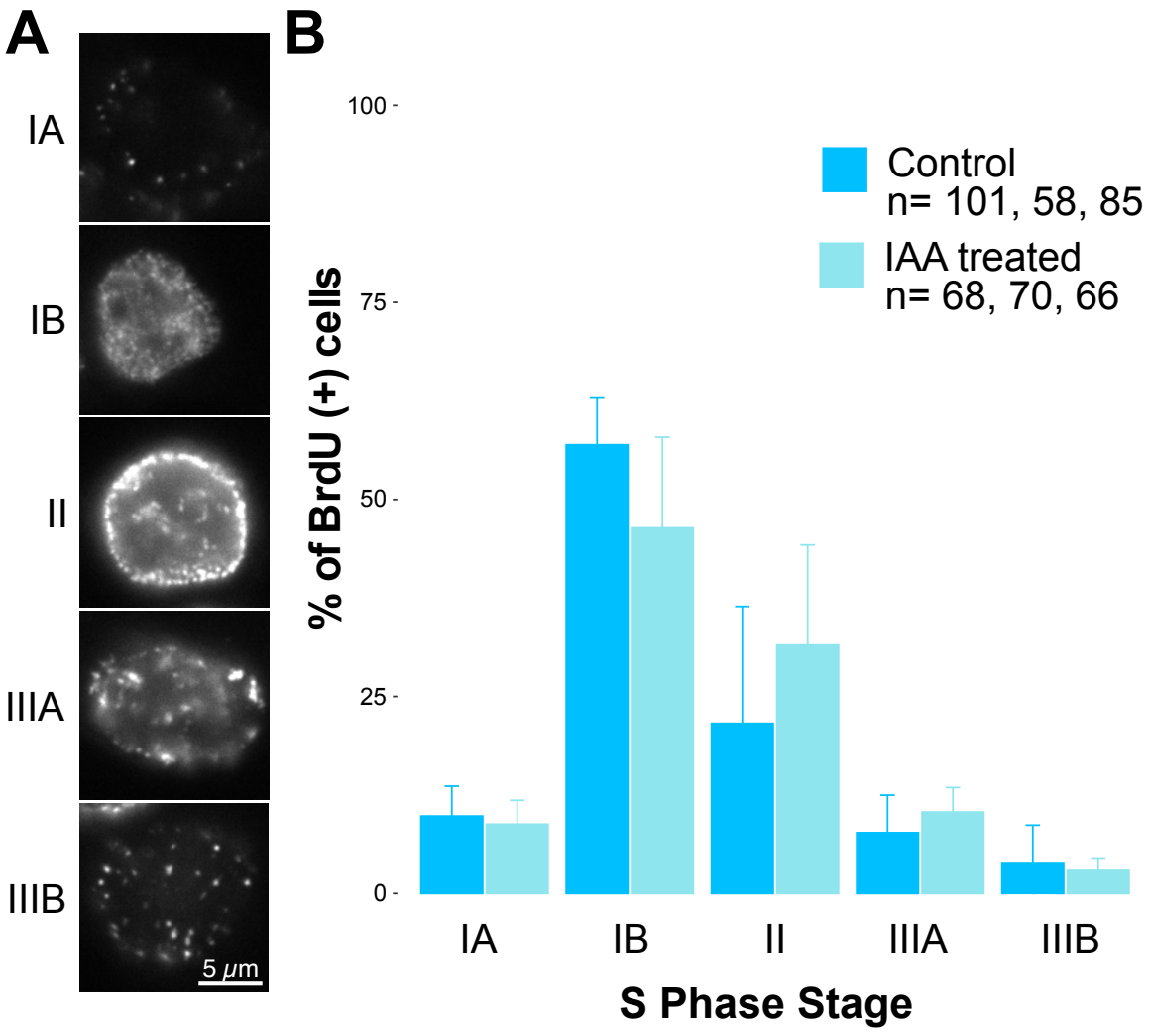


**Figure 14**. **Acute loss of cohesin in S phase does not perturb patterns of replication foci**. **(A)** Replication foci patterns were assayed using BrdU IF in asynchronous cells after auxin or control treatment. **(B)** Mean and 90% confidence interval for the proportion of cells in each S phase stage based on BrdU foci classifications. Three biological repeat experiments were carried out, the total number of BrdU-positive cells analyzed from each repeat listed as (n).

### *Depletion of SCC1 from early G1*

Replication timing is established early in G1, in a period described as the timing decision point (TDP) [113]. Thus, to assess whether cohesin is required for the establishment, rather than execution, of replication timing, it was necessary to ablate cohesin from cells before the onset of S phase. This also diminished the potential for artifacts arising from the impact of cohesin ablation on chromosome segregation. Cells were synchronized with a 24 hour thymidine block and released for 3 hours, followed by nocodazole treatment for 8 hours. Cells were released from nocodazole by mitotic shake-off and 30 minutes after the shake-off, auxin was introduced for treated cells. Flow cytometry was used to assess the synchrony upon release and the SCC1 depletion in the auxin-treated sample [Figure 15]. In both the control and treated arms, cells progressed through S phase with tight synchrony [Figure 15B]. However, a subset of cells from both the control and auxin-treated conditions were found to remain in G1 without incorporating BrdU after the mitotic shake-off [Figure 15A]. This can be anticipated based on the fact that the thymidine-nocodazole synchronization regimen is likely to cause some cell stress. The non-cycling cells were determined not to be a problem for the experiment as the lack of BrdU incorporation means that the genomic material of these cells is excluded in the IP step.

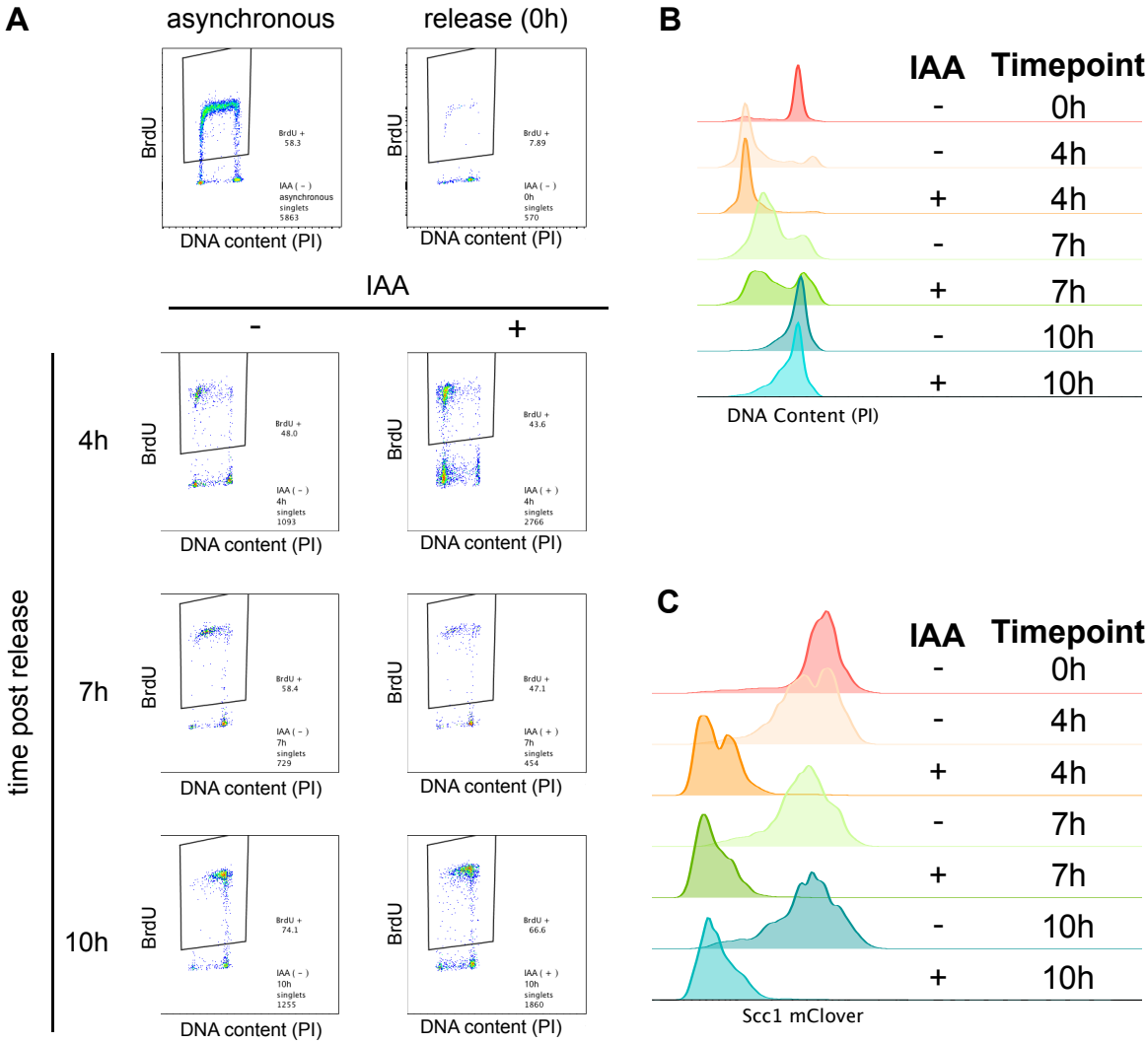



**Figure 15. Cell cycle synchronization for synchronized Repli-Seq.** Cells were synchronized with a 24 hour thymidine block followed by a 3 hour release into an 8 hour nocodazole block. Cells were released from nocodazole by mitotic shake-off and treated with auxin or kept as controls 30 minutes after the shake-off. Cells were harvested at 4, 7, and 10 hours following release and were pulsed with BrdU 2 hours prior to harvest (*i.e.* at 2, 5, and 8 hours, respectively). At the point of harvest for Repli-Seq cells were also harvested for flow cytometry. **(A)** Flow cytometry scatterplots showing DNA content (PI) versus BrdU signal. The BrdU positive gate is shown, and the percentage of BrdU positive cells is listed for each sample. The 'asynchronous' sample comes from a different experiment where asynchronous cells were pulsed with BrdU, and is shown as a reference. **(B)** Density plots of the DNA content of the BrdU-positive cells from each timepoint. **(C)** SCC1 depletion in the auxin-treated sample was confirmed by assessing mClover signal with flow cytometry.

Synchronized auxin-treated or untreated control cells were pulsed with BrdU at 2, 5, and 8 hours post-release. 2 hours after the addition of BrdU to each synchronized timepoint (*i.e.* at 4, 7, and 10 hours) the cells were harvested and DNA was extracted. BrdU-labeled nascent DNA was immunoprecipitated and

subjected to high-throughput sequencing. The overall replication timing profiles for synchronized cells with cohesin depleted from early G1 showed the same pattern to those of untreated synchronized cells [Figure 16A]. The variability between treated and control cells was comparable to that between control cells and a published independent control HCT116 dataset [31] [Figure 16C]. 219 bins showed significant changes in response to cohesin ablation (out of 56,034 50-kb bins with map score > 0.5), and these were spread across the genome (median distance between bins 700,000 bases) and enriched proximal to unmappable regions such as centromeres [Figure 16B]. Just 14 of these bins were shared with the significant bins from the asynchronous sample, and in four of these bins the direction of the change was different – suggestive of noisy or unmappable loci rather than a consistent biological effect [Table 1].

**Table 1**. Genomic loci showing significant difference in both asynchronous and synchronized experiments.

| | | | RT Control - RT IAA | |
|---|---|---|---|---|
| Chr | Start | End | Asynchronous | Synchronous |
| chr3 | 198100000 | 198150000 | -3.440337 | 3.242949 |
| chr4 | 49300000 | 49350000 | 2.723062 | 2.296154 |
| chr4 | 119400000 | 119450000 | 3.486406 | -1.410933 |
| chr4 | 190000000 | 190050000 | 4.479181 | -1.897292 |
| chr4 | 190050000 | 190100000 | 12.678844 | -5.136778 |
| chr4 | 190100000 | 190150000 | 25.976415 | -10.545368 |
| chr5 | 49650000 | 49700000 | -3.258203 | 4.272367 |
| chr5 | 181250000 | 181300000 | 2.921178 | -1.623559 |
| chr7 | 56850000 | 56900000 | 2.779265 | 1.355865 |
| chr9 | 150000 | 200000 | 3.343623 | -1.447844 |
| chr13 | 114300000 | 114350000 | 3.460791 | 1.561016 |
| chr13 | 114350000 | 114364328 | 7.856382 | 3.064328 |
| chr16 | 90100000 | 90150000 | 4.566427 | -3.221925 |
| chr21 | 6350000 | 6400000 | 3.521525 | -2.910266 |

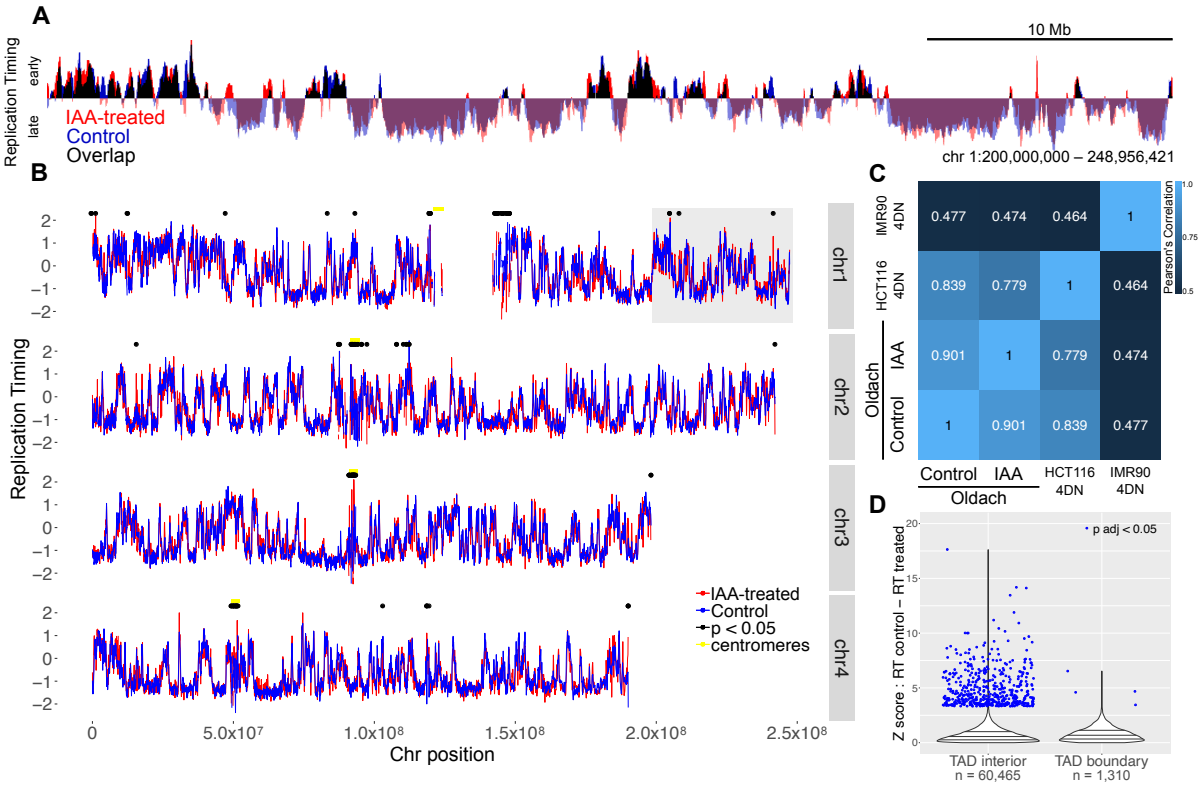



**Figure 16**. **Loss of cohesin from early G1 does not perturb replication timing**. **(A)** Example locus overlay of replication timing (RT, log-scaled ratio of 4 hour timepoint read counts over 10 hour timepoint counts) for auxin-treated versus control cells. **(B)** RT plots for four example chromosomes with significantly changed loci marked with black dots. The gray box highlights the region shown in (A). **(C)** Pearson correlation calculated for RT values across genome, compared to previously published 4D Nucleome Repli-Seq data. **(D)** Violin plot showing the distribution of Z scores for control versus auxin-treated RT difference across the genome, separating out genomic bins covering a TAD boundary.

As with the replication timing data from asynchronous cells, no specific genomic features showed a particular change in replication timing [Figure 17].

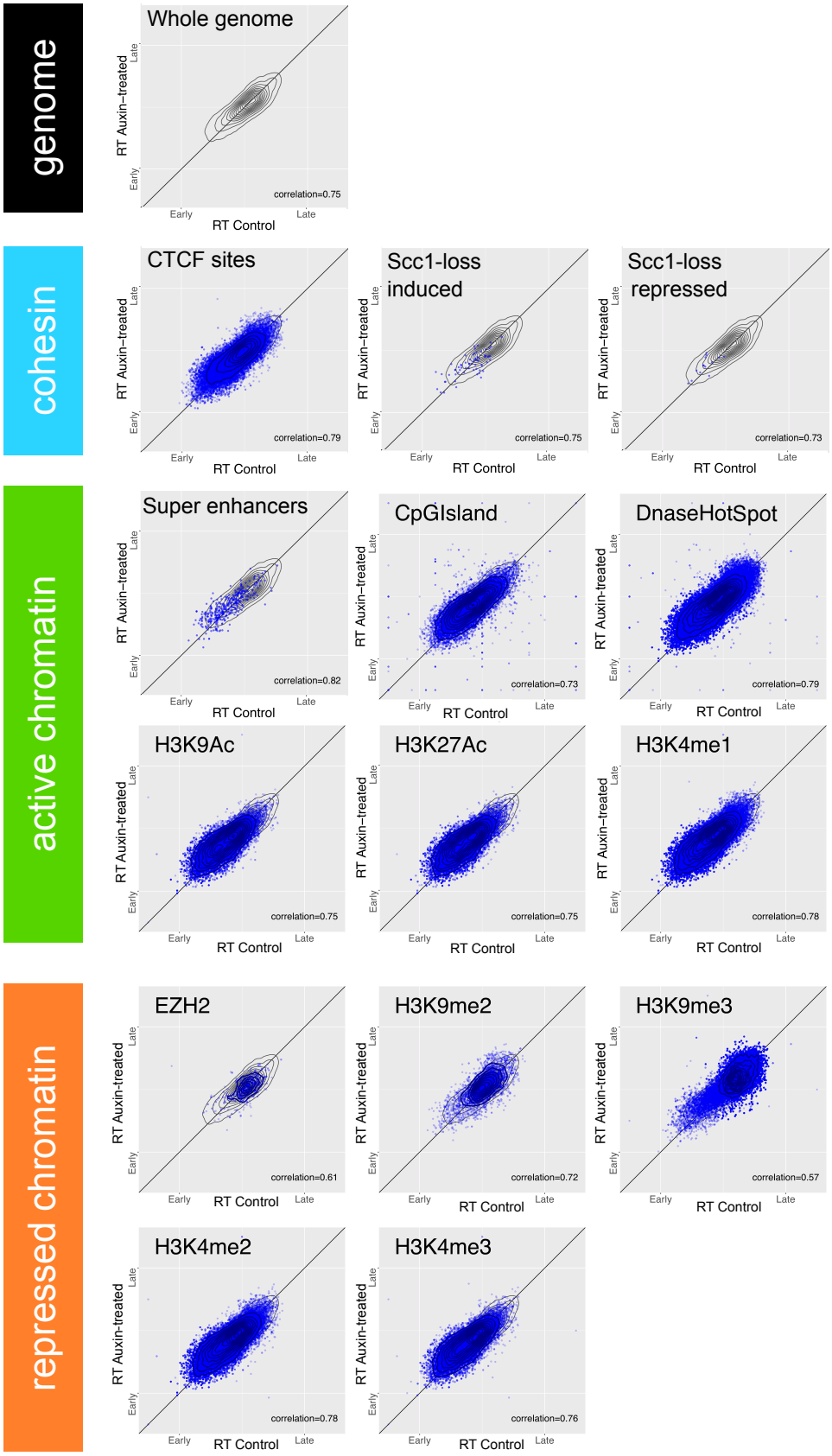



**Figure 17. Correlation between control and auxin-treated synchronized RT in genome regions associated with cohesin or replication timing**. For each bin in the genome, RT can be compared between the control and treated samples. Genome regions covering CTCF binding sites and the set of genes upregulated in response to cohesin-ablation (SCC1-loss induced) or

downregulated in response to cohesin-ablation (SCC1-loss repressed) [34] were specifically assessed for changes in RT in response to cohesin ablation. Genome regions with histone marks and DNA binding sites associated with early replication timing (active chromatin) or late replication timing (repressed chromatin) were also assessed. The external datasets used to define genome regions are listed in Materials and Methods. The whole-genome behavior is plotted as a grey topology for reference. The dark blue points and topology indicate the RT for genomic bins annotated with for each characteristic or genome feature.

The cytological patterns of replication foci were again analyzed to determine whether replication was organized differently in the nucleus in the absence of cohesin from early G1. As cells were not adherent for microscopy following mitotic shake-off, ImageStream flow-microscopy was used for cell visualization [114].

ImageStream analysis software was used to identify cells with BrdU stain patterns matching training sets of cells chosen as representative of the patterns of 'small foci', 'peripheral ring', or 'large foci,' corresponding to the patterns seen in early, mid, and late S phase, respectively [Figure 18]. As expected, the majority of cells classified as showing a BrdU pattern of 'small foci' were in the early (4 hour) sample, and most cells of the 'big foci' class were in the late (10 hour) sample. Both control and cohesin-depleted samples showed similar numbers of cells for each class in each timepoint sample. The only significant difference between auxin-treated and control samples was seen in counts of 'big foci' cells in the 4 hour timepoint ($p$ = 0.03, an average of 27% of the 4 hour cells being classed as having 'big foci' in the auxin-treated sample versus 11% in the control). This lack of significant differences between replication foci patterns following auxin-induced cohesin ablation suggests that cohesin loss neither alters the patterns of replication timing at a genome sequence level, or at a genome organizational level.

This was a preliminary attempt to investigate whether ImageStream might be feasible as a robust approach to high-throughput replication foci classification.

These initial findings based on four classification runs were promising, but it would be valuable to carry out further work to rigorously establish the validity of this method. Further development of this approach in the future could allow for the assessment of orders of magnitude more cells while reducing the risk of human bias.

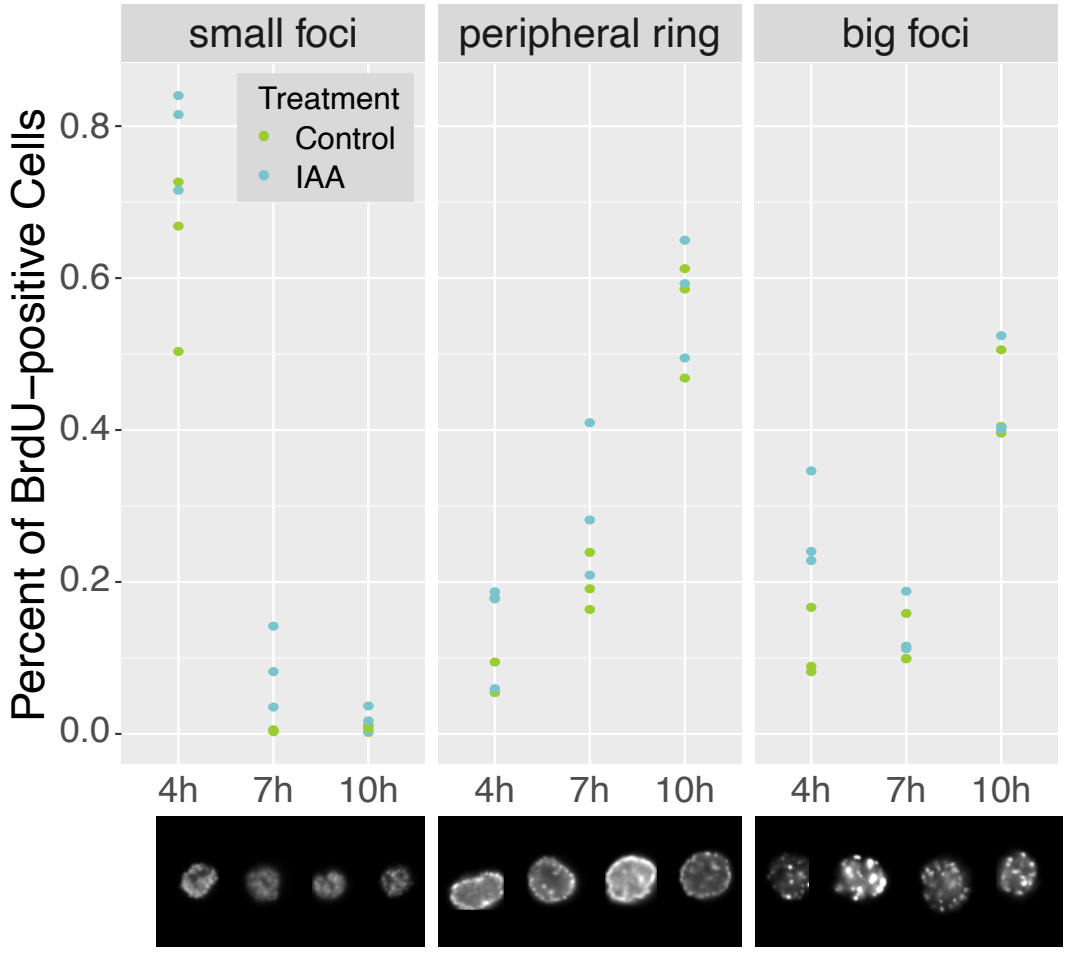



**Figure 18**. **ImageStream analysis of BrdU foci**. Cohesin ablated or control cells were pulsed with BrdU at 2, 5, and 8 hours following release from synchronization and auxin treatment and were harvested at the end of the 2 hour pulse. Cells were imaged using ImageStream. Three independent classification runs were carried out using distinct training cell sets on the ImageStream software to identify cells matching the replication patterns of 'small foci,' 'peripheral ring,' or 'large foci' across each of the treated or control samples. T tests were run to compare the auxin-treated or control cells for each time point and foci class: of each of the tests only the 4 hour sample for the 'big foci' classifier was significant ($p < 0.05$). (Bottom) Example images of cells exhibiting each of the BrdU foci patterns.

Impact of genome organization on replication timing

The lack of response of replication timing domains to cohesin ablation led to the hypothesis that compartment boundaries may be a more robust delineator of replication timing than TAD boundaries. To assess this possibility, TAD calls from the control and auxin-treated cells of Rao *et al*, 2017 were collected from the 3D Genome Browser [37]. Upon cohesin ablation, Rao *et al* saw a pervasive loss of TAD structure while higher level compartment structure remained [Figure 19A] [34]. In keeping with this, of 2,026 compartment-independent TAD boundaries (*i.e.* not within 500 kb of a compartment boundary) mapped in the untreated HCT116 cells, only 887 remained following auxin treatment. TAD boundaries that disappeared in the auxin-treated dataset were classed as 'SCC1-dependent boundaries' (n=1,139), while those that were shared (within 500 kb) between control and auxin-treated datasets were classed as 'invariant' (n=887), and any that fell within 500 kb of an A/B compartment domain boundary were analyzed separately (n=1,378).

Replication timing values for auxin-treated and control cells were assessed along 10 Mb windows to either side of SCC1-invariant and dependent boundaries. Both control and auxin-treated datasets showed a similar relationship between replication timing domains and TAD boundaries: the loss of TAD organization did not lead to changes in replication timing across TAD boundaries [Figure 19B]. Furthermore, SCC1-invariant and compartment boundaries were found to show a stronger relationship to replication timing domains than SCC1-dependent boundaries, as evidenced by a sharper change in replication timing at the boundary [Figure 19B].

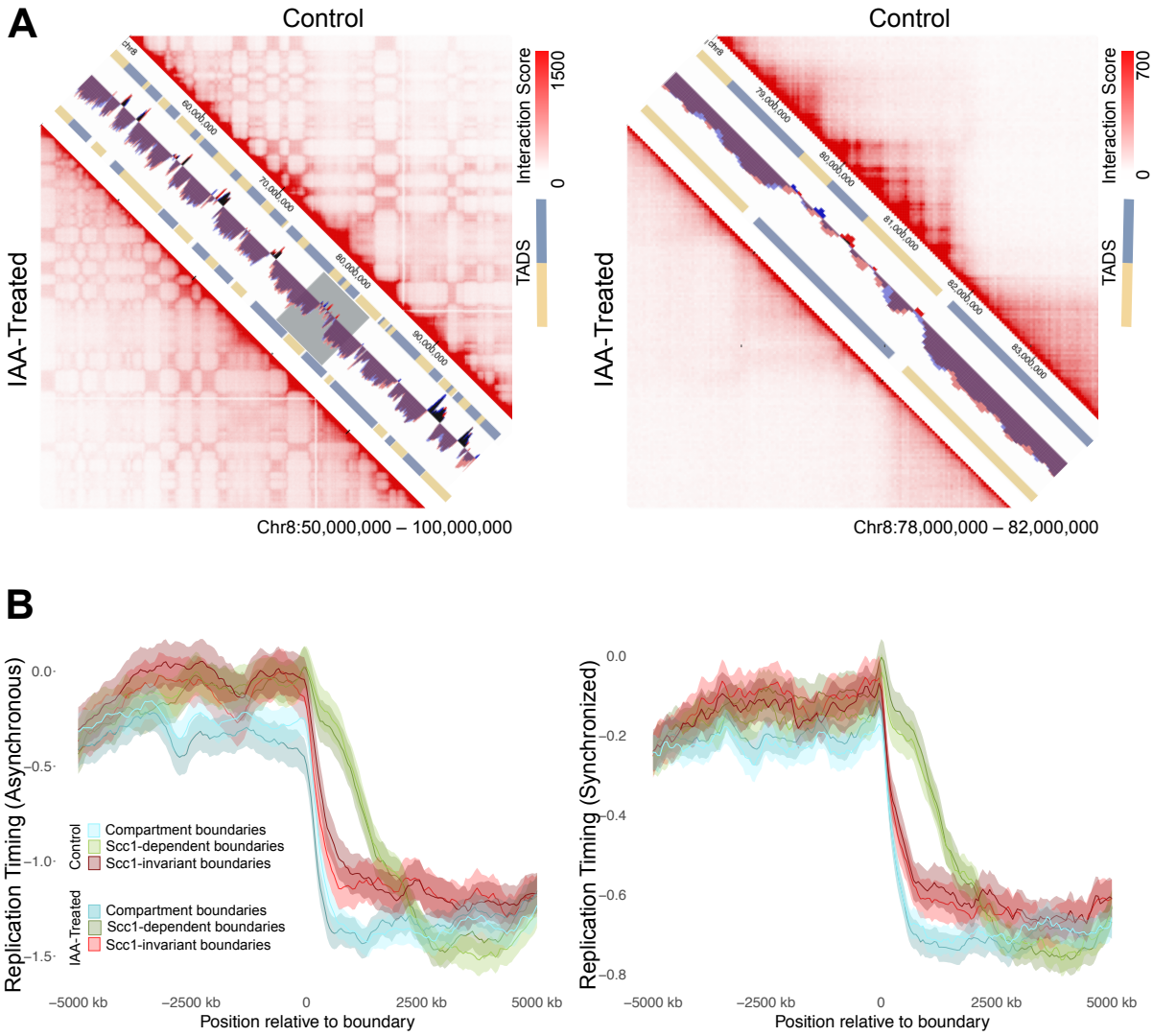



**Figure 19**. **Replication timing across TAD boundaries versus compartment boundaries. (A)** Visualization of HiC data from Rao *et al*., 2017 in untreated cells versus auxin-treated cells. The left panel spans 50 Mb to highlight compartments and the right zooms in to 6 Mb from within that locus (grey box) to highlight changes to TAD structure. **(B)** Replication timing metaplots for windows 5 Mb on either side of TAD boundaries. Invariant boundaries are those called in both control and auxin-treated cells, SCC1-dependent boundaries are those which were lost upon auxin treatment [34].

## Discussion

The hypothesis that TADs act as a fundamental unit of organization for replication timing has gained widespread support due to the near one-to-one mapping of replication timing domains and TADs [58], the concurrent establishment of both domains in early G1 [77], and their coordinated changes over the course of cellular differentiation [66,78]. Using the SCC1-degron system, this work was able to explicitly test whether the loss of cohesin triggered a perturbation to replication timing patterns. It resulted that neither acute loss of cohesin during replication in S phase nor prolonged loss of cohesin from early G1 drove significant changes to replication timing patterns. Thus, cohesin is concluded not to be required for the execution nor establishment of replication timing domains. This is in good agreement with recent work showing that depletion of TAD boundary protein CTCF does not result in significant genome-wide perturbations to replication timing patterns, nor do specific TAD boundary changes via CTCF binding site ablations result in localized replication timing changes [115].

While Repli-Seq can yield deep insight into replication timing domains, the ensemble nature of the approach limits its power for identifying specific origin usage. Thus, it remains feasible that cohesin-dependent structure may play a role in the specific sites of origin firing within a domain, while the overall timing of higher-order domains remains independent of TAD structure. This could reconcile the apparent lack of domain timing change to the changes in interorigin distance seen previously in cohesin knockdown experimentation [7]. Single-cell Repli-Seq or a single-molecule approach to map DNA replication such as DNA combing or DNAscent [44] have the potential to deconvolute population data to look specifically

at initiation sites utilized in a single cell or on a single chromosome, and thus may yield important insights into origin selection and cooperativity at play within a given replication timing domain.

The lack of response of replication timing domains to cohesin loss and the stronger relationship seen between cohesin-invariant and compartment boundaries than cohesin-dependent boundaries to changes in replication timing suggests that replication timing domains may be regulated by the same organizing principles that drive A/B compartmentalization rather than TADs. While this could in part be explained by the fact that neighboring TADs can share similar chromatin states, and thus may be likely to replicate at similar times regardless of being members of distinct regulatory modules, there is increasing evidence to support a model in which compartmentalization-based structure rather than loop-based structure drives replication timing. Recently-discovered early replication control elements (ERCE) have been shown to drive robust CTCF-independent interactions in a manner akin to compartmentalization, and their deletion leads to changes in replication timing [115]. Further work characterizing the cellular mechanisms driving genome compartmentalization will open pathways to perturb this system and elucidate the causal relationship between replication timing and genome topology.

## Perspective and Further Work

Following the completion and publication of this project [52], more work has been published by other groups to understand the role of cohesin-mediated chromosome architecture in genome function, and specifically the role of chromosome architecture in DNA replication.

The finding, described herein, that replication timing domains persist despite cohesin ablation was replicated and validated in a subsequent large-scale study by Cremer *et al.* employing Repli-Seq approaches along with HiC and microscopy in the SCC1 auxin degron system [116]. The authors confirmed that loss of cohesin did not drive significant changes in replication timing profiles or replication domain boundaries.

The Cremer study recapitulated findings observed herein that cohesin depletion did not impact cell progression through S phase and G2. Interestingly, the authors pushed forward to follow cells through mitosis in the absence of cohesin. Unsurprisingly, mitotic progression was found to be markedly delayed and aberrant, as expected given the critical role of cohesin in chromosome segregation. The cohesin-depleted post-mitotic cells failed to form two daughter cells, and instead remained as single cells with multilobulated nuclei (MLN). Amazingly, despite the aberrant 2n-nuclei and complete loss of TAD structure, HiC and microscopy on the cohesin-ablated post-mitotic MLN showed a persistence of A/B compartments and cytological patterns of active/inactive nuclear compartments (including chromosome territories, interchromatin channel system, and the organization of chromatin with different compaction levels). Furthermore, *de novo* DNA synthesis in post-mitotic MLN after 30 hours of auxin-induced cohesin ablation still showed the characteristic cytological replication foci patterns expected for early, mid, and late S phase.

Cremer *et al.* further extended understanding of the impact of cohesin on DNA replication with detailed analysis of replication foci captured with super resolution microscopy. The authors found replication domains to be subtly increased in

volume and number in the cohesin-ablated cells, suggesting a role for cohesin in compacting and cohering replication domains, if not driving their formation.

Together, the Cremer study and this work show that the maintenance of replication domains and the temporal program of DNA replication is independent of cohesin-mediated loop architecture. While rejecting the role for cohesin-mediated genome architecture in replication, both this study and Cremer *et al.* re-emphasize the strength of genome compartmentalization as a predictor of spatiotemporal patterns of replication timing. Thus, the hypothesis that the genome's spatial organization critically drives replication timing is not rejected. However, instead of a mechanism of loop-mediated origin contact allowing for facilitated passage of limiting replication initiating factors, a mechanism that now seems more likely is that phase separation drives the creation of subnuclear microenvironments with distinct limiting factor concentrations [Figure 20].

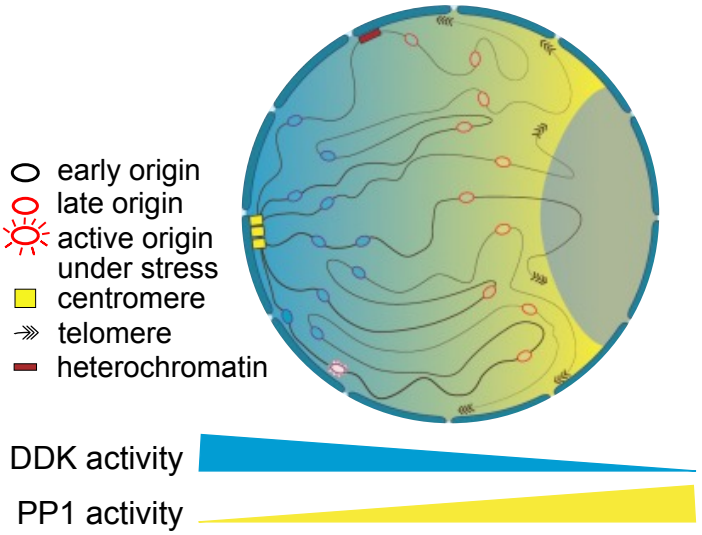



**Figure 20. A/B compartments form microenvironments with distinct availabilities of replication initiation factors.** Figure from [75].

Testing this mechanism is challenged by the fact that specific drivers of genome-wide compartmentalization have been hard to pin down. Instead,

compartmentalization is likely achieved through complex interactions of a wide array of genome characteristics and binding partners undergoing phase separation [72,73]. The complexity and robustness of the compartmental-level genome organization (surviving even in the aberrant, 2n multilobed nuclei seen in Cremer *et al.*) pose substantial challenges to studies hoping to perturb compartments genome-wide in order to show impacts on replication timing. Furthermore, nuclear compartmentalization has been suggested to be critical not only for the DNA replication program, but for many facets of genome function – most notably, transcription [117,118]. Thus, there may be a careful balance to be struck in finding techniques that perturb compartmentalization genome-wide while preserving cell health sufficiently to draw physiologically relevant conclusions.

The finding by Sima *et al*. of early replicating control elements (ERCEs) that drive both compartment identity and replication timing in a CTCF-independent manner thus marks not only a substantial contribution to the field, but also a valuable opportunity to push forward the features that drive compartmentalization as well as its link to replication timing. The three original ERCEs identified by Sima *et al.* share three major characteristics: an enrichment of active chromatin marks, binding of transcription factors important for cell identity, and CTCF-independent interactions.  On the basis of these characteristics, 1,835 ERCEs have been putatively identified in the mouse genome, and a further three ERCEs validated [115].

Further work will be required to identify the effector proteins and mechanisms that confer the interactions and early replication timing unifying ERCEs. Among the candidates are the bromo-domain containing proteins BRD2 and BRD4, which

bind acetylated histones known to be enriched in the ERCEs [119]. The BRD2 and BRD4 proteins have recently been identified as binding Treslin (Sld3 in *S. cerevisiae*), one of the rate-limiting initiation factors recruited to the MCM helicase following DDK-mediated phosphorylation, opening an intriguing possibility for BRD2/BRD4 as ERCE effector proteins [119]. Once identified, the *trans* factors acting on ERCEs will likely join Rif1 and Forkhead as significant drivers of both genome organization and replication timing. These factors may reveal further mechanisms for the targeted recruitment of replication initiation proteins to specific genomic compartments and may also shed light on the broader question of genome-wide drivers of compartmentalization.

To test the overarching mechanistic hypothesis – that genome compartments form microenvironments that replicate with distinct timing due to varying local concentrations of limiting initiation factors [Figure 20] – methods to either perturb compartmentalization or to study limiting factor localization and diffusion will be important. For the latter, the advance of high-resolution single-molecule tracking microscopy studies will likely be invaluable. These studies have been successfully carried out in live-cell mammalian systems to observe the chromatin interaction dynamics of a variety of nuclear proteins [120] and could be applied to DDK or other limiting factors to assess whether distinct gradients of localized concentrations exist in the nucleus during S phase, and whether these change as S phase progresses.

# Chapter IV. Investigation of polymerase epsilon exchange at the *S. cerevisiae* replication fork

# Introduction

## Eukaryotic genome replication

The replication of genomic DNA proceeds from bidirectional replication forks emanating outward from an origin. For each fork, the template DNA is unwound by a helicase complex into two antiparallel strands. Each of these strands must be replicated to allow for the successful division of the original cell into progeny with complete genomes.

Importantly, for a eukaryotic replication fork moving down a given length of DNA, the two template strands pose very different substrates for replication, as one is unwound in a 5' to 3' direction, and the other 3' to 5'. As DNA polymerases synthesize only in a 5' to 3' direction, this necessitates two distinct mechanisms for replication of the two strands. The 3' to 5' template – or leading strand – can directly track behind the unwinding helicase, potentially synthesizing the new strand of DNA in a processive manner. The 5' to 3' strand – the lagging strand – must be copied in the opposite direction from that of helicase unwinding, resulting in discontinuous synthesis producing a gapped new strand composed of Okazaki fragments [16].

In eukaryotic cells, there are three essential DNA polymerases, alpha, epsilon, and delta, which carry out the bulk of genomic DNA replication [121]. Polymerase alpha, uniquely lacking a 3' to 5' proofreading exonuclease domain, performs an initiating role in starting DNA synthesis off of the RNA primer synthesized by a primase. This leaves the two remaining polymerases, epsilon and delta, to carry out the bulk of genomic DNA replication.

Much work has focused on elucidating and differentiating the contributions of polymerase epsilon and delta to DNA replication. This has largely supported a 'division of labor' hypothesis, which suggests that these two polymerases each specialize in the distinct replication mechanisms required for leading and lagging strand synthesis [122,123]. Early after its characterization in *S. cerevisiae*, the relatively high processivity of the polymerase epsilon catalytic subunit, *POL2,* over that of polymerase delta, *POL3,* raised conjectures for polymerase epsilon's involvement in leading strand synthesis [121]. While this hypothesis has been iteratively refined, the most widely accepted model in *S. cerevisiae* suggests that lagging strand synthesis is carried out by polymerase delta and the majority of leading strand synthesis is carried out by polymerase epsilon [18,122,124,125]. Adding nuance to the division of labor model, recent work has highlighted a role for polymerase delta on the leading strand specifically at sites of initiation and termination [18,122,124,125]. Furthermore, evidence is mounting that polymerase delta contributes to proofreading on both the leading and lagging strands [126].

Polymerase epsilon has been implicated in important roles in replication outside of its direct role in leading strand synthesis. Among these are replisome maturation, modulation of replisome progression, and coordination of the S phase checkpoint. This work aims to further investigate the role that polymerase epsilon plays in replication by assessing its stability at the replisome.

In particular, this work addresses the question of whether a single Pol2 subunit of the polymerase epsilon complex remains functionally associated with a replisome from replication initiation to termination, or whether there can be dynamic exchange of the Pol2 subunit over the course of a single replicon. Elucidating

whether Pol2 is stably associated with a moving replisome will not only advance understanding of the dynamics and coordination of replisome machinery, but also has fundamental implications for genome stability and the ability of the replisome to cope with replication fork impediments.

## Polymerase epsilon structure and regulation

In *S. cerevisiae,* polymerase epsilon is a heterotetramer of subunits Pol2, Dpb2, Dpb3, and Dpb4 existing in 1:1:1:1 stoichiometry [127,128] [Figure 21]. The largest of its subunits is the catalytic subunit, Pol2, a bilobed protein of around 200 kDa [121].

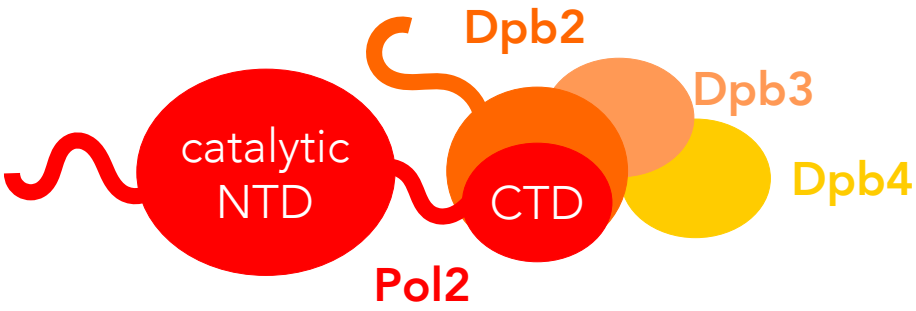



**Figure 21. Polymerase epsilon complex schematic**. Adapted from [129].

The synthetic function of the polymerase epsilon complex is housed in the Pol2 subunit: *in vitro* primer extension reactions showed similar synthesis rates for the Pol2 subunit as for the complete polymerase epsilon complex [130]. Interestingly, Pol2 encodes two polymerase and two 3' to 5' exonuclease domains. The non-catalytic C terminus contains an inactive version of the polymerase and exonuclease, while the N terminus features enzymatically-active copies of these two domains, along with a domain that acts to switch the DNA between the synthetic and proofreading modes [131]. However, while yeast can survive a truncation of the catalytically-active N terminus of Pol2, the non-catalytic Pol2 C terminus has been shown to be essential for yeast viability [Table 2] [121]. This

essential terminus plays an important structural role in driving polymerase epsilon holoenzyme formation.

**Table 2.** Selected polymerase epsilon mutants highlighting Pol2 function.

| Strain | Mutation | Phenotype | Study |
|---|---|---|---|
| *pol2-9* | M643I (polymerase domain) | DNA replication and cell growth temperature sensitivity. | [132] [discussed in [133]] |
| *pol2-18* | P710S (polymerase domain) | DNA replication and cell growth temperature sensitivity. | [132] [discussed in [133]] |
| *pol2-M644G* | M644G (polymerase domain) | 'Mutator variant': drives promiscuity of nucleotide incorporation, proofreading not compromised. | [125,126] |
| *pol2-X / pol2-DD875,877AA* | D875A + D877A (polymerase domain) | Lethal. Destroys Pol2 polymerase activity but the presence of the polymerase domain blocks other polymerases from compensating. | [133,134] |
| *pol2-D640A* | D640A (polymerase domain) | Lethal. Destroys Pol2 polymerase activity but the presence of the polymerase domain blocks other polymerases from compensating. | [135] [134] |
| *pol2-11* | Nonsense (30 amino acid truncation) | DNA replication temperature sensitivity and slow growth at the permissive temperature. Cell S/M checkpoint defects. | [136–138] [discussed in [133]] |
| *pol2-12* | Nonsense (26 amino acid truncation) | DNA replication temperature sensitivity and slow growth at the permissive temperature. Cell S/M checkpoint defects. | [136,137] [discussed in [133]] |
| *pol2-16* | Deletion of the catalytic domains | Viable and supports DNA replication, but | [18,138–140] |

| | | | |
|---|---|---|---|
| | (AA 1-175…1135-2222) | with slow growth and prolonged S phase | |
| *pol2-Δcat* | C terminal truncation (AA 1,263-2,222) | Can support initiation and both leading and pol delta-dependent lagging strand synthesis *in vitro.* | [141,134] |
| *pol2-4* | D290A and E292A (exonuclease domain) | No exonuclease activity *in vitro*, increased mutation rate *in vivo.* | [18,142,143] [discussed in [131]] |
| *pol2-REL* | R567C, E611K, L621F (N terminus but outside of catalytic domains. Copying mutations found in cancers.) | Temperature sensitive. Defective for pre-LC formation, although polymerase epsilon complex formation fine. Asymmetric replication structures. Replication checkpoint intact. | [144] |
| *pol2-KR / pol2-K571R* | K571R | Abolishes Pol2 SUMOylation in HU-treated cells, and basal SUMOylation observed in unperturbed S phase. | [145,146] |

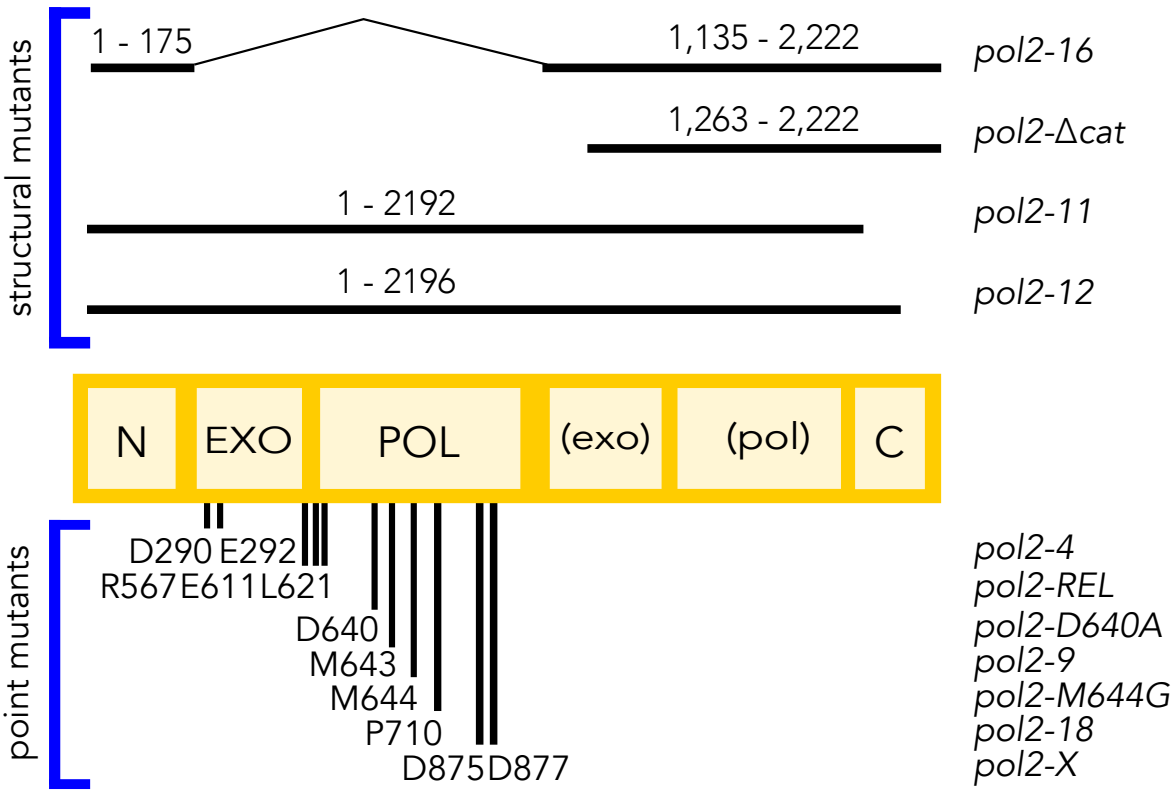



**Figure 22**. **Pol2 protein domains and point and structural mutations of interest.** The Pol2 protein is composed of two domains connected by a flexible linker. The catalytic N terminal domain (N) houses an exonuclease (EXO) and polymerase (POL) domain. The C terminal domain is non-catalytic and houses an inactive exonuclease (exo) and inactive polymerase (pol) domain. Various point and structural mutants [Table 2] have been developed to map the Pol2 functions in replicative synthesis, proofreading, and CMG maturation; their mutation points are mapped above and below the Pol2 schematic. Adapted from [134,147,148].

Of the three smaller subunits, only the 80 kDa Dpb2 protein is essential for yeast growth. Dpb2 has been shown to be integral for polymerase epsilon to form a complex with the CMG [11]. In keeping with this, the temperature sensitive *dpb2-1* mutant was shown to be defective for yeast chromosome replication [149].

The final two factors, Dpb3 and Dpb4, are small ancillary proteins that, while not essential for cell viability, have been shown to be important for efficient and precise DNA replication. The *dpb3Δ* mutant impedes polymerase epsilon stability and confers a slight mutator phenotype [150]. Dpb4 disruption similarly results in diminished polymerase epsilon stability, and *dpb4Δ* mutants show delayed cell

cycle dynamics resulting from slow progression though S phase [151]. While their role in replication has not been fully characterized, the predicted structure of Dpb3 and Dpb4 includes a histone fold motif seen in other proteins associated with modifying chromatin structure [151].

Polymerase epsilon expression and dynamics

The regulation of replisome component abundance could play an important role in protecting against genome replication undertaken outside of S phase or under S phase conditions unconducive to faithful genome duplication. This regulation can be achieved through two approaches: control of replisome component expression or programmed degradation. It appears that both approaches may be at play in the case of DNA polymerase complexes.

The *POL2* gene shows periodic expression patterns with a peak at the G1/S boundary [Figure 23] [132,152,153]. Polymerase epsilon subunits *DPB2* and *DPB3* share expression regulatory sequences, and *DPB2* shares a similar cell-cycle dependent expression pattern to that seen in *POL2* [149,150].

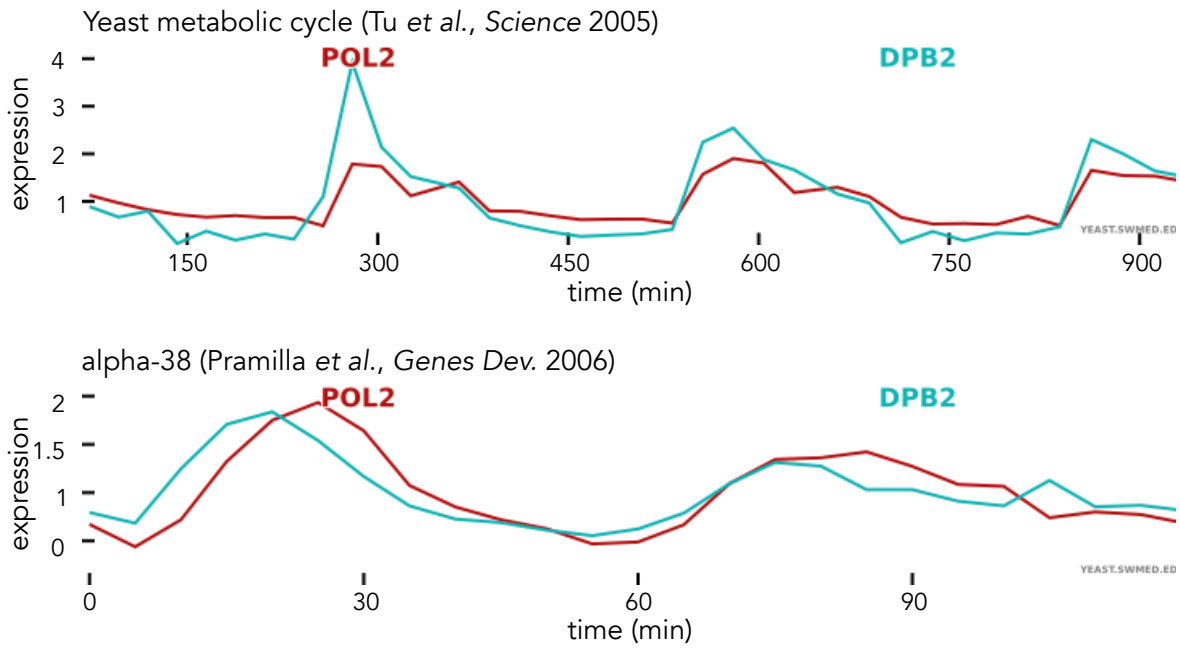



**Figure 23**. **Cell cycle-dependent mRNA expression patterns in polymerase epsilon subunits *POL2* and *DPB2*.** Data visualization from: [154].

Mechanisms for the coordinated degradation of polymerase epsilon subunits have not been fully elucidated in *S. cerevisiae*, but a few lines of evidence hint at their potential importance. One of these lines of evidence includes a system that has been characterized in *S. pombe* in which polymerase epsilon and delta subunits, Pol2 and Pol3, and MCM subunits undergo coordinated $SCF^{Pof3}$-dependent degradation in conditions of replication stress in the absence of fork protection by Swi1 [155,156].

In contrast to this global degradation of leading and lagging replisome components under replication stress, in *S. cerevisiae* Pol3 has been shown to be specifically targeted for proteasomal degradation in response to DNA damage [157]. The specific degradation of the polymerase delta catalytic subunit has been suggested as an elegant mechanism for maintaining replisome progression while allowing for lesion bypass via a switch to a translesion synthesis (TLS) polymerase upon encountering specific template obstructions to replication.

An analogous mechanism for polymerase epsilon has not been characterized, but recent publications have described Pol2 sumoylation in response to HU [145,146]. This post-translational modification is suggested to have a role in allowing polymerase epsilon progression to adapt in response to various replication challenges. In *S. pombe,* Pol2 protein was found to be markedly less stable than Pol3 even in cells without any specific replication stress and with uncompromised fork protection complex [155]. If this holds in *S. cerevisiae,* a high level of Pol2 turnover could be a method for clearing post-translational modifications such as the Pol2 sumoylation once stressed conditions are corrected.

## Polymerase epsilon roles in S phase

While the classification of Pol2 as a DNA polymerase immediately suggested its likely function in DNA replication, work has shown that it is not strictly essential for genome duplication either *in vivo* [138] or *in vitro* [141]. This raises the question of what roles the polymerase epsilon complex plays in an unperturbed S phase – or why redundancy in DNA polymerases has been maintained throughout eukaryotic evolution.

### *CMG (Cdc45 – MCM – GINS) helicase maturation*

While polymerase epsilon is dispensable to DNA synthesis – albeit at a cost to replication speed and fidelity, as discussed below – the non-catalytic C terminus of Pol2 and the Dpb2 subunit are essential for cell viability. This has been attributed in part to the role of polymerase epsilon in CMG maturation and maintenance.

Muramatsu *et al.* first showed in 2010 that polymerase epsilon forms a fragile complex with Dpb11, Sld2, and GINS, termed the pre-loading complex (pre-LC).

The components of the pre-LC assembled independent of DNA and were found to be mutually dependent on one another for origin association [9]. This suggested a model in which polymerase epsilon, Dpb11, Sld2, and GINS load onto DNA only after complexation into the pre-LC. This hypothesis was supported by *in vitro* replication work, which showed the recruitment of GINS to MCM to be dependent on the full complement of polymerase epsilon, Dpb11 and Sld2 [8].

Further characterization of the pre-LC mapped the interaction to GINS specifically to an N terminal domain of Dpb2 that bound the Psf1 subunit of GINS [11,158]. *In vitro* CMG assembly studies showed that neither Pol2 nor Dpb2 alone were sufficient to enable CMG formation – both essential subunits of polymerase epsilon are required for CMG maturation and thus play a critical role in replication initiation [158,159]. In keeping with the *in vivo* evidence that Dbp3 and Dbp4 are not essential, CMG assembly *in vitro* was found to be as efficient with these factors as it was without them [158].

A conserved region in the Pol2 C terminus outside of the catalytic domain has also been implicated in CMG assembly [144]. Mutation of three conserved residues of this region, yielding a mutant named *pol2-REL,* resulted in defects in pre-LC formation and reduced association of CMG subunits Cdc45 and Psf1 with MCM in S phase cells [144]. There was still some replication in this mutant, however, and associations between Dpb11, GINS, and Sld2 were observed in the absence of polymerase epsilon, leading the authors to suggest that a ‘three member’ complex may be sufficient – although less effective – for delivery of GINS to MCM in the absence of competent polymerase epsilon [144].

### *CMG progression and replisome speed*

Although catalytically-dead mutants of polymerase epsilon are sufficient for CMG maturation and other polymerases can compensate for DNA synthesis, polymerase epsilon is still required for maximal replisome speed.

The first evidence for this came from observations of *S. cerevisiae* strains lacking in polymerase epsilon catalytic activity. While these strains were viable, they exhibited cell cycle defects resultant from slow S phase progression, along with tiny spore colonies despite larger than normal cell size, and aberrant nuclei [18,138]. Interestingly, the tethering of leading strand synthesis to CMG emerged as an important factor for replisome speed. In *S. cerevisiae* strains expressing an N terminal truncation of Dpb2 (which supports CMG assembly and maintains polymerase epsilon catalytic activity, but abrogates tethering of the polymerase to the CMG), immunoprecipitation experiments showed functional and fully assembled replisomes with all canonical components except for polymerase epsilon, but cell proliferation was severely delayed [11].

*In vitro* reconstituted replication reactions corroborated the *in vivo* observed requirement for polymerase epsilon (along with Cms3/Tof1/Mrc1) for maximum fork speeds. Although polymerase delta was sufficient for replication of both leading and lagging strand templates, the reaction proceeded 3 to 10-fold slower than when both polymerase epsilon and delta were present [141,162]. Similarly, leading strand products produced in reconstituted replication reactions with catalytic-null polymerase epsilon were shorter than those produced in reactions with wildtype polymerase epsilon. This is consistent with the model of less-processive synthesis supported by polymerase delta in the absence of

polymerase epsilon catalytic activity [158]. Again, the specific tethering of polymerase epsilon to the CMG emerged as an important factor. Langston *et al.* separately purified a three-subunit polymerase epsilon and Dpb2, and assayed the production of full length product after replication of a CMG-bound template [159]. While the Pol2-Dpb3-Dpb4 complex alone was inefficient for full length product formation in the specified time window, increasing concentrations of Dpb2 rescued the synthesis [159].

*In vitro* systems have also been employed to quantify the extent of the slowing. Using microscopy to study helicase progression past leading or lagging strand lesions on single molecules *in vitro*, Sparks *et al.* found that the *Xenopus* egg extract CMG maintained its progression rate of around 400-500 nucleotides per minute after bypassing a lesion on the lagging strand, but slowed to around 80 nucleotides per minute after bypass of a lesion on the leading strand [163]. The authors hypothesized that this slowing was due to uncoupling from leading strand synthesis, which was supported by the finding that in the absence of lesions, aphidicolon slowed the CMG to an average speed of 55 nucleotides per minute [163]. Strikingly similar rates have been recorded from *in vitro* reactions reconstituted with *S. cerevisiae* replication proteins: when uncoupled due to limiting dNTP conditions, the CMG was calculated to unwind at a rate of around 55 bases per minute, whereas the same reconstituted system under non-limiting dNTP conditions showed synthesis at 1 – 2 kilobases per minute [134]. *In vivo* studies have utilized the catalytically-null *pol2-16* mutant to study fork progression and DNA synthesis rates in a system lacking physical coupling between the CMG and leading strand synthesis. By employing ChIP for PCNA progression from the early-firing *ARS305* origin at 16°C, Ohya *et al* estimated the rate of fork

progression to drop from 366 nucleotides per minute for wildtype cells to 132 nucleotides per minute for *pol2-16* mutants [140]. Assessed by the introduction of a heavy nucleotide, the rate of DNA synthesis was found to drop from 1,380 nucleotides per minute for wildtype cells to 540 nucleotides per minute for *pol2-16* cells at 25°C [140].

The complete set of determinants of CMG speed have not yet been fully elucidated, but a solved structure of the *S. cerevisiae* CMG in complex with polymerase epsilon may shed light on a mechanism through which polymerase epsilon could modulate CMG progression [158]. An interaction between the Pol2 C terminus and Dpb2 was found to change how the helicase contacted a DNA fork. In previous MCM structures solved with inactive ATPase motors, a winged helix appendix of Mcm5 blocked the MCM central channel – potentially blocking DNA binding [164]. However, in complex with polymerase epsilon, the Pol2 C terminus contacted the Mcm5 C-terminal winged helix, which was hypothesized to stabilize the aperture of the MCM pore, allowing for DNA engagement [158]. In addition to the interaction at the winged helix, the C terminus of Pol2 was found to bridge the gap between Mcm5 to Mcm2 ATPase gate, while Dpb2 bridged the Mcm5-3 ATPase gate [158]. This bridging was suggested to be effective in stabilizing the MCM ring planar conformation, and thus loss of polymerase coupling at a barrier could contribute to CMG pausing, while sustained polymerase binding keeps the CMG in a conformation conducive to maximal speed.

The structural observations do not yet prove causation, and some *in vitro* work suggests that CMG deceleration upon uncoupling may not be polymerase specific, but rather responsive to the loss of leading strand synthesis in general.

The CMG slowed at a leading strand lesion even in a reconstituted system with a catalytically-dead polymerase epsilon, where polymerase delta carried out leading strand synthesis [165]. Further work to fully clarify the mechanism of CMG progression will be important to elucidate exactly what role polymerase coupling plays. Irrespective of the exact mechanism, this responsiveness of the helicase to leading strand synthesis critically allows for a balance to be struck in the extent of single stranded DNA exposed. Sufficient unwinding is required to allow for repriming or the activation of the replication checkpoint, but extensive unwinding can deleteriously expose large tracts of damage-vulnerable single stranded DNA. Thus, the ability of the CMG to sense and respond to associated synthesis is critical to maintaining genome stability.

### *Replisome fidelity*

Beyond replisome speed, polymerase choice can also have implications on replication fidelity. *S. cerevisiae pol2-16* mutants lacking catalytic domains were found to rapidly acquire suppressor mutations [18], and to have higher mutation rates than wildtype yeast [166]. One study found mutagenesis to halve upon deletion of the gene encoding the polymerase zeta catalytic subunit, with the persistent mutations corresponding to the well-characterized polymerase delta mutation spectrum [166]. This suggests that in the absence of polymerase epsilon catalytic capacity *in vivo,* synthesis was carried out by polymerase zeta as well as polymerase delta, resulting in increased mutation rates and diminished proofreading.

*pol2-REL*, a mutant with three N terminal residues outside of the Pol2 catalytic domain mutated, shed further light on the potential importance of polymerase

epsilon in maintaining genome fidelity. Instead of a marked increase in point mutations, this mutant showed an increased propensity for gross chromosomal rearrangement (GCR). At the permissive temperature these were increased 6-fold, and at the higher semi-permissive temperature GCRs were observed at rates increased 350-fold over wildtype [144]. An increase in GCR rates was also seen in the checkpoint-defective *pol2-11* C terminal truncation mutant [167].

While a mechanism for this increase has not be definitively characterized, it is interesting to note that the *pol2-REL* mutant showed synthetic lethality with accessory helicases and exonucleases required to deal with replication obstacles such as DNA-bound proteins or R loops [144]. This leads to the hypothesis that changes to the non-catalytic Pol2 C terminal domain could lead to a defect in the ability of polymerase epsilon to overcome template barriers, resulting in an increase in replication stalls and subsequent double strand breaks and rearrangements.

### *Replication checkpoint involvement*

After successful initiation of replication, continued coordination between the helicase and polymerase throughout genome replication is important to prevent the exposure of extensive tracts of damage-vulnerable single stranded DNA. However, this coordination can be challenged by factors that prevent the continued progression of the polymerase, such as a lesion on the leading strand template or a paucity of the nucleotides needed for new strand synthesis. In these cases, polymerase-helicase uncoupling results in an accumulation of single stranded DNA, which is sensed by the apical checkpoint sensor kinase, Mec1, leading to the activation of transducer kinase Mrc1, and subsequently the effector

kinase Rad53 [168,169]. A cascade of protective measures follow, including stabilization of stalled forks, upregulation of nucleotide synthesis and DNA repair genes [170], inhibition of initiation from late replication origins [171], and cell cycle arrest [172].

Hydroxyurea (HU) inhibits the ribonucleotide reductase required for new nucleotide synthesis, and thus HU treatment is a well-established method for inducing replication stress and eliciting the replication checkpoint. Early work testing *pol2* mutants' sensitivity to HU hinted at the important role of polymerase epsilon in stimulating checkpoint activation [133]. Specifically, the role in checkpoint activation was mapped to the C terminus of Pol2, as C terminal truncation mutants *pol2-11* and *pol2-12* failed to activate the checkpoint in response to HU and died as a consequence of cells entering mitosis with unreplicated DNA [137]. In contrast to the finding of failed checkpoint activation in *pol2-11* and *pol2-12* truncation mutants, Garbacz *et al.* found that the N-truncated *pol2-16* strain demonstrated a constitutively activated replication checkpoint signature. This suggests that leading strand synthesis by polymerase delta is unable to fully compensate for the role that polymerase epsilon plays in checkpoint mediation [166].

One potential difference in checkpoint activity may stem from the distinct polymerase responses to nucleotide scarcity. In *rad53* mutants, HU treatment resulted in the exposure of long tracts of single stranded DNA, as expected given the replication checkpoint deficiency of these mutants [173]. Interestingly, DNA synthesis under HU stress in these mutants progressed farther on the lagging strand than the leading strand – suggesting that leading strand synthesis by polymerase epsilon may be particularly sensitive to low dNTP levels *in vivo* [173].

This was further supported by *in vitro* reconstituted replication on a chromatinized template [173].

This differential sensitivity could be driven by distinct proclivities for template switching between the polymerase and exonuclease domains; *in vitro* polymerase epsilon has been shown to favor exonuclease digestion in conditions of low dNTP concentration [130]. In other B family polymerases, the switch between exonuclease and polymerase domains is governed by an extended β-hairpin loop absent from the Pol2 structure, suggesting that an alternative structural feature tunes Pol2 between synthesis and excision [174]. In addition to intrinsic differences, the identification of a Pol2 site (K571) specifically sumoylated in response to HU treatment suggests a potential active signaling mechanism through which polymerase epsilon may respond to low dNTP levels [145,146].

The *pol2-16* mutant lacks both the catalytic domain and the specific sumoylation site but houses an intact C terminal domain. Intriguingly, despite sufficient origin loading to support replication initiation, a single-locus *in vivo* IP analysis found that Pol2-16 protein did not demonstrate the same stability with the replisome in response to HU treatment as that seen for wildtype Pol2 [140]. While the Pol2 protein was found at the *ARS306* region upon release from G1 and was retained in the region over a period of incubation with HU, Pol2-16 was loaded normally but quickly lost from the region during HU arrest, despite the presence of RPA and PCNA being unchanged in the *pol2-16* versus the *POL2* strain [140]. This single locus finding merits further study on a wider genomic context but suggests a role of either the catalytic domains or perhaps the sumoylation site in stabilizing

polymerase at the fork in response to replication stress, important for allowing forks to continue when conditions have improved.

Polymerase epsilon is also distinct from other replicative polymerases in its interactions with a variety of proteins involved in the checkpoint response. Some of these interactions which may play a role in shaping checkpoint activation or effector functions are described below.

### $RFC^{Ctf18}$

Replication factor C-like (RFC) complexes function to load or unload DNA clamps such as PCNA, which can then act as scaffolding for the recruitment of factors important for replication functions such as lesion bypass, fork recovery, or chromatin inheritance [175]. One RFC variant, $RFC^{Ctf18}$, has been shown to be a critical component for both the intra-S phase replication stress checkpoint and the establishment of sister chromatid cohesion. $RFC^{Ctf18}$ forms a stable complex with polymerase epsilon [176], binding on the outer surface of the exonuclease domain of Pol2 [147]. The strength of this binding has led to the suggestion that $RFC^{Ctf18}$ may track along with polymerase epsilon, remaining in position to load PCNA onto the leading strand to contribute to checkpoint activation rapidly in response to fork stalls [147]. Indeed, variants of the $RFC^{Ctf18}$ complex that disrupt polymerase epsilon binding have shown decreased viability and reduced Rad53 activation in HU [175–177]. Human $RFC^{Ctf18}$ has been shown to stimulate polymerase eta activity – in keeping with the known function of clamps in acting as scaffolding for polymerase recruitment and retention – suggesting a potential role for $RFC^{Ctf18}$ in the pathway of activating lesion-bypass polymerases in response to stalling at template obstacles [178].

## Mrc1

Mrc1 (mediator of the replication checkpoint) was identified as a required factor for Rad53 activation in response to replication stress [172]. In addition to its important role in checkpoint signal transduction, Mrc1 has also been shown to have effector functions directly on the replisome. Mrc1 associates with the replisome upon origin activation and travels with the fork [179]. It has been shown to be required for maximum fork speeds in reconstituted replication reactions [180], and to be critical for helicase stalling in response to HU treatment [179].

Immunoprecipitation experiments have shown direct association of Pol2 and Mrc1. In asynchronous cells, about half of the cellular pool of Mrc1 was pulled down with Pol2, while almost all Pol2 molecules were associated with Mrc1 [139]. In unperturbed S phase, this Pol2-Mrc1 association is driven by independent interactions of Mrc1 to both the N and C termini of Pol2 [139]. When exposed to DNA replication stress, Mrc1 is phosphorylated, which interrupts the Pol2 catalytic N terminal interaction but leaves the C terminal interaction strong [139].

This raises the possibility that Mrc1 could contribute to driving a switch between the two Pol2 conformations, ensuring that when no replication stress is sensed, Pol2 is stabilized in a conformation promoting engagement with the DNA, but when the checkpoint is activated, the Pol2 catalytic terminus is released. In potential support of this notion of Mrc1 playing a supportive role in normal replication, Mrc1-null mutants show slowed progression through S phase even in the absence of exogenous replication stress [172,181].

Can *et al.* mapped out the role that Cdc45, along with Mrc1, plays in recruiting Rad53 to the replisome [182]. Intriguingly, both of these proteins, which are required

for targeting Rad53 to replication forks, have multiple points of contact with polymerase epsilon. In Cdc45, one polymerase epsilon binding site (discussed in more detail later) in particular is directly adjacent to the Cdc45 loop region involved in Rad53 binding [Figure 24] [182]. Thus, this hints at further potential involvement of polymerase epsilon in the replication checkpoint through dynamic interplay with these proteins. The direction of the interaction remains unclear: is the synthesis of the leading strand polymerase sensed to drive replication checkpoint activation, or is polymerase synthesis directly modulated in response to checkpoint activation? Further work to characterize these interactions and their implications will be useful in building understanding of how the replication checkpoint functions to protect genome integrity.

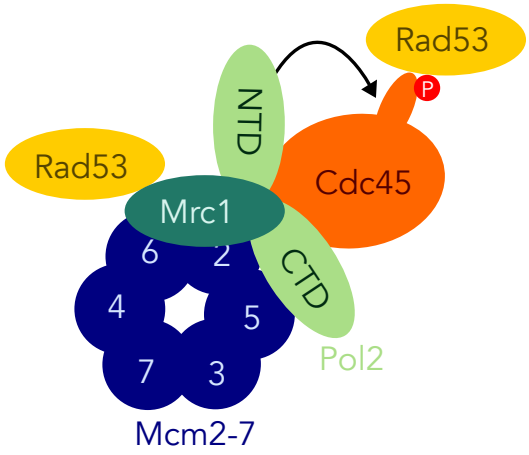



**Figure 24. Binding interactions bridging the CMG, Pol2, and checkpoint mediators**. Rad53 binds Cdc45 and Mrc1 – aiding its recruitment to the replisome – and both interact with Mcm2-7 and polymerase epsilon. One polymerase epsilon conformation forms an interaction between the Pol2 catalytic N terminal domain (NTD) and the region of Cdc45 that drives Rad53 binding. This suggests a mechanism by which the synthetic status of polymerase epsilon could be relayed to replication checkpoint activation. Adapted from [182].

### *Replication-coupled nucleosome assembly*

For complete heritability of genomic information, chromatin status must be replicated along with DNA. Thus, the replication process requires first the successful removal of nucleosomes from template duplex DNA (to allow

replisome progression), and then the recycling of the epigenetically marked parental histones onto appropriate DNA loci. Parental histones are supplemented with additional newly synthesized histones to accommodate the doubled DNA load. This replication-coupled nucleosome assembly in budding yeast is largely coordinated by histone chaperone Asf1, which brings newly synthesized H3-H4 dimers to the fork, and CAF-1, recruited by PCNA, which deposits histones onto the newly replicated DNA [183]. Based on the histone-fold motif in the structure of Dpb3 and Dpb4, and the finding that polymerase epsilon co-purified with histones, it was suspected that this polymerase might also play a role in chromatin removal or reassembly during replication [184].

Using enrichment and sequencing of protein-associated nascent DNA (eSPAN), Yu and Gan *et al* found that the parental histones – which importantly have the post-translational marks for appropriate expression level – displayed a bias towards the lagging strand, such that at least 23% more parental histones were loaded onto lagging strands [185]. The newly synthesized histones (identified by H3K56 acetylation) showed a compensatory bias for the leading strand [185]. Surprisingly, knockout of Dpb3 or Dpb4 exaggerated this bias; *dpb3Δ* and *dpb4Δ* strains showed the same overall level of chromatin-bound parental and new histones, but showed 120% more parental histone loaded on the lagging strand than the leading strand [185]. This suggested a role for Dpb3 and Dpb4 subunits in ensuring symmetrical loading of the recycled parental histones onto the leading as well as lagging strands.

The functional importance of strand-symmetrical histone reassembly becomes apparent in the propagation of silenced loci. Both *dpb3Δ* and *dpb4Δ* yeast strains

showed silencing defects at telomeric genes and the mating locus, with the Dpb3 knockout conferring a slightly greater loss of silencing than Dpb4 [184,186]. Similarly, an *S. cerevisiae* strain encoding a Dpb3 histone fold mutant with diminished DNA binding capacity was found to lose telomeric silencing with greater frequency than wildtype yeast [187]. While the Pol2 catalytic mutant (*pol2-11*) also showed partial loss of telomeric gene silencing, this could be rescued with overexpression of Dpb3 and Dpb4 [186]. Thus, polymerase epsilon subunits facilitate epigenetic inheritance via coordination of replication-coupled nucleosome assembly, in addition to the polymerase's clear role in replicating DNA for genetic inheritance.

Response to lesions

The differences in synthesis tactics for the two unwound template strands of DNA hold strikingly different implications for responding to replication obstacles. The inherent discontinuity of lagging strand synthesis allows for a straightforward bypass of a polymerase delta-impeding lesion; synthesis can simply carry on at the next newly-primed Okazaki fragment [188]. This leaves a limited single stranded gap, which can be efficiently filled in by a translesion polymerase. Critically, this response depends upon efficient repriming, which is already an important part of canonical lagging strand synthesis.

On the leading strand, however, the tactic of replicating over the lesion with gap filling synthesis is challenged by the fact that leading strand restart is considerably more rare [188]. *In vitro* replication studies of a cyclobutene pyrimidine dimer (CPD) showed polymerase alpha-dependent repriming to be the rate-limiting step for leading strand replication restart [188]. However, more recent work suggested that lesion bypass by translesion polymerase eta frequently works via an 'on the fly'

model, avoiding replication restart and gap filling and instead relying on a switch from the replicative DNA polymerase to the specialized translesion enzyme [189].

While many questions about the nature and regulation of the switch remain, the model built from *in vitro* replication reconstituted with a lesion on the leading strand and translesion polymerases eta in addition to canonical polymerases alpha, epsilon, delta, suggests that polymerase epsilon releases the nascent strand after stalling and the CMG progresses, uncoupled [189]. It is implied that polymerase epsilon stays with the CMG, as seen in previous studies [173]. When the reaction was spiked with a leading-strand primer that binds downstream of the lesion, replication restart was efficient and resulted in two discontiguous leading strand products. In the absence of a restart primer, polymerase eta was competent for replicative bypass of the lesion and a single full-length leading strand product was generated. High polymerase delta concentrations inhibited lesion bypass, but this inhibition was lost upon monoubiquitination of PCNA. This suggests that polymerase delta can outcompete polymerase eta for the nascent strand, but that PCNA ubiquitination acts as a regulatory mechanism for the specific recruitment of a translesion polymerase. Interestingly, in the complete absence of polymerase delta, while polymerase eta was highly efficient at lesion bypass, a minority of the products reached full length and instead a smear of products of lower molecular weight were seen. This suggested a requirement for polymerase delta to reclaim the nascent leading strand downstream of the lesion to allow for recoupling to the CMG and polymerase handoff back to polymerase epsilon. This suggestively mirrors the requirement for polymerase delta at replication initiation [18], and hints at a possible deficiency of polymerase epsilon in capturing nascent ends from the primase or translesion polymerases. Further

study into the nascent end exchange from polymerase delta to polymerase epsilon will be very interesting to uncover both the method by which this is carried out and the potential reasons for polymerase epsilon's dependency upon polymerase delta in these replication transition periods.

Dynamic processivity

Some of the first intimations that polymerase epsilon acts in leading strand synthesis came from observations of its processivity, which suggested its suitability for tracking along behind the unwinding helicase [121]. Interestingly, in *in vitro* work on single stranded, pre-primed templates in the absence of CMG or PCNA, polymerase epsilon was shown to have a higher affinity for DNA than polymerase delta, but only a slightly higher processivity [190]. Upon addition of PCNA (but still without CMG), the processivity of polymerase delta was substantially increased (100-fold), while the processivity of polymerase epsilon was augmented only moderately (6-fold) [190].

This raised the questions: what characteristics of polymerase epsilon account for its increased affinity for DNA, and what additional factors are at play *in vivo* to confer polymerase epsilon its higher observed processivity than polymerase delta? The first solved structure of the catalytic core of the Pol2 subunit of polymerase epsilon in complex with DNA suggested an answer to the first question. The authors identified a domain unique amongst all B family polymerases (of which *S. cerevisiae* polymerase alpha and delta are members) that allowed Pol2 to encircle nascent double stranded DNA [174]. They termed this unique domain the 'Processivity' or 'P' domain, and found that mutations in three residues of the P domain abrogated the polymerase's processivity [174].

Later structural work expanded the outlook to the full polymerase epsilon in complex with the CMG [Figure 25]. This revealed the polymerase epsilon complex to consist of two flexibly tethered lobes, with the catalytic domain of Pol2 making up one of the lobes and the noncatalytic C terminus – joined by the noncatalytic polymerase epsilon subunits – another [127]. The noncatalytic lobe is anchored to the ATPase face of the CMG through Dpb2 and Pol2 C terminal interactions to GINS, Cdc45, and MCM. Even when in complex with the CMG, polymerase epsilon retained its flexibility; the catalytic lobe was found to in two distinct conformations, one in which it was bent inward towards the helicase ring and another in which it projected outward from Cdc45 [127].

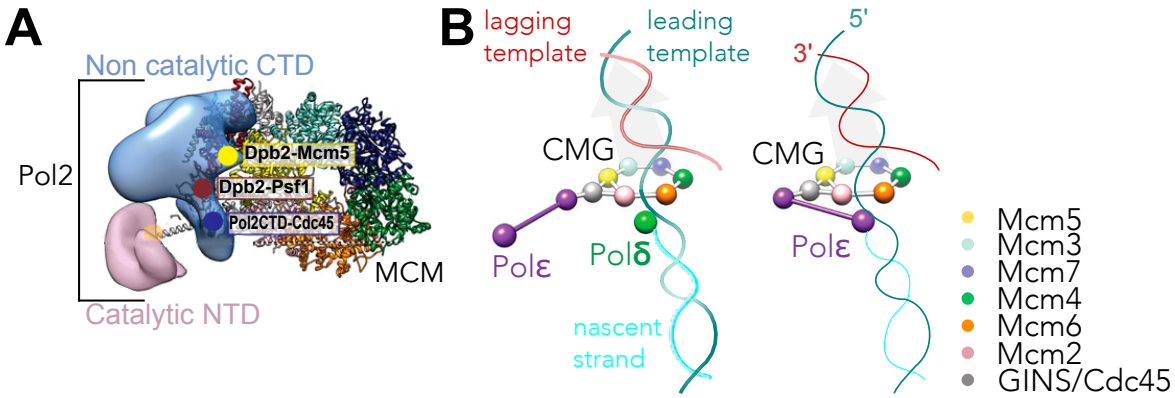



**Figure 25. (A)** Electron microscopy structure of polymerase epsilon in complex with the CMG shows its bi-lobed form, with the noncatalytic lobe including the C terminus of Pol2, Dpb2, Dpb3, and Dpb4 bound to the ATPase face of the CMG. The catalytic N terminus of Pol2 is attached by a flexible linker and was crystallized in two distinct conformations. **(B)** Schematic of polymerase-switch mechanism for the establishment of leading strand synthesis. The reconfiguration of Pol ε catalytic domain allows for engagement of Pol ε with the leading strand template after initial synthesis by Pol δ. Both **(A)** and **(B)** adapted from [127].

These two conformations suggested possible functional states: one in which the polymerase epsilon catalytic domain is engaged with DNA and actively synthesizing or proofreading, and one in which the polymerase epsilon catalytic domain is removed from the DNA template while still engaged with CMG. This model critically allows for decoupling of polymerase epsilon catalytic activity from its presence and functioning at the fork, which is especially compelling given the

known requirement for polymerase handoff events – from polymerase alpha via polymerase delta after initiation [18,141] and to translesion polymerases at obstacles. This model of transient dissociation from the primer and template but continued association with the CMG has been termed 'Dynamic Processivity' [139,191].

Polymerase epsilon exchange

The model for dynamic processivity suggests a mechanism by which polymerase epsilon complexation to the CMG could remain entirely static over the duration of S phase despite dynamic engagement of the polymerase with the DNA. However, the extent to which this stability is actually observed at the replisome – or whether there are conditions which favor the exchange of polymerase epsilon subunits – remain open questions.

An *in vitro* co-immunoprecipitation assay of *S. cerevisiae* replication components suggested that exchange can occur. This assay addressed switching of Cdc45 and polymerase epsilon subunit Dpb2 between CMG complexes by mixing S phase extract from three cell lines expressing either (1) Tap-tagged Mcm3, Myc-tagged Dpb2 and Myc-tagged Cdc45, (2) Flag-tagged Cdc45 or (3) Flag-tagged Dpb2 [11]. The CMG pulled down by Tap IP included exclusively Myc-tagged Cdc45 and not Flag-tagged Cdc45, indicating the high stability of Cdc45's incorporation into the CMG such that *in vitro* switching of Cdc45-Mcm3 binding partners did not occur in the mixed S phase extracts [11]. In contrast, the Tap IP pulled down both Myc- and Flag-tagged Dpb2, giving *in vitro* evidence for the ability of polymerase epsilon subunits to switch into CMG complexes following assembly [11].

This experiment does not shed light, however, on whether this switching can occur in the context of a progressive replisome which has already initiated

replication, or the frequency with which this switching might occur. More pertinent to this question have been microscopy studies, which enable the observation of single replisomes and can yield insight into their dynamic behaviors over time.

Early microscopy studies were carried out on bacterial cells, which allowed *in vivo* analysis of ongoing replication due to the relative simplicity of the entire genome being replicated by two forks emanating outward from one origin. One such *E. coli* fluorescence recovery after photobleaching (FRAP) and single particle tracking (SPT) study found the majority of replisome proteins – with the exception of the DnaB helicase – to show high turnover at the fork [192]. The observed disengagement from the replisome and exchange of the polymerase subunit suggested both leading strand and lagging strand synthesis to be carried out discontinuously [192]. Another study similarly showed rapid exchange of the bacterial polymerase subassembly during DNA replication, but found that in the absence of free polymerase in solution the replisome was able to carry on – suggesting the potential for stable binding of the polymerase to the helicase, but an inclination towards exchange when possible [193]. This finding is mirrored in work showing that the addition of catalytically dead polymerase to an ongoing *in vitro* replication reaction resulted in the inhibition of replication, suggesting an ability of free polymerases to compete out polymerases assembled into an ongoing fork [194].

It is not possible to extrapolate directly from the bacterial replisome to the eukaryotic replisome, but these microscopy studies significantly advanced understanding of the dynamic workings of the replisome, and opened avenues of enquiry into how the eukaryotic replisome might compare.

Using a tethered-bead assay Lewis *et al.* were able to visualize *in vitro S. cerevisiae* reconstituted leading strand synthesis in real time [180]. The authors found that when the CMG helicase had been assembled in complex with polymerase epsilon on DNA, replication did not depend upon continued presence of additional polymerase epsilon in the flow buffer, suggesting that, like in the bacterial system, polymerase epsilon binding to the CMG could be stably maintained. However, in the presence of competitor polymerases in solution, *S. cerevisiae* polymerase alpha, delta, and epsilon all showed rapid concentration-dependent exchange, while CMG remained stably bound irrespective of its concentration in solution [195].

Åberg, Duderstadt, and van Oijen suggested that a multisite competitive exchange mechanism can reconcile the apparent incongruity of a highly stable molecular machine like the replisome exhibiting rapid and frequent exchange of specific subunits [196]. In this mechanism, an abundance of low-affinity binding sites – such as the multiple interactions characterized between polymerase epsilon, the CMG, and the additional replisome accessory factors – allows for the binding of excess polymerases. Thus, a competitor polymerase could be held in proximity to the actively engaged polymerase, and upon transient dissociation of the active polymerase from an effector binding site the complex can proceed either with re-binding of the original polymerase or with new binding of the competitive polymerase, leading to polymerase exchange [Figure 26].

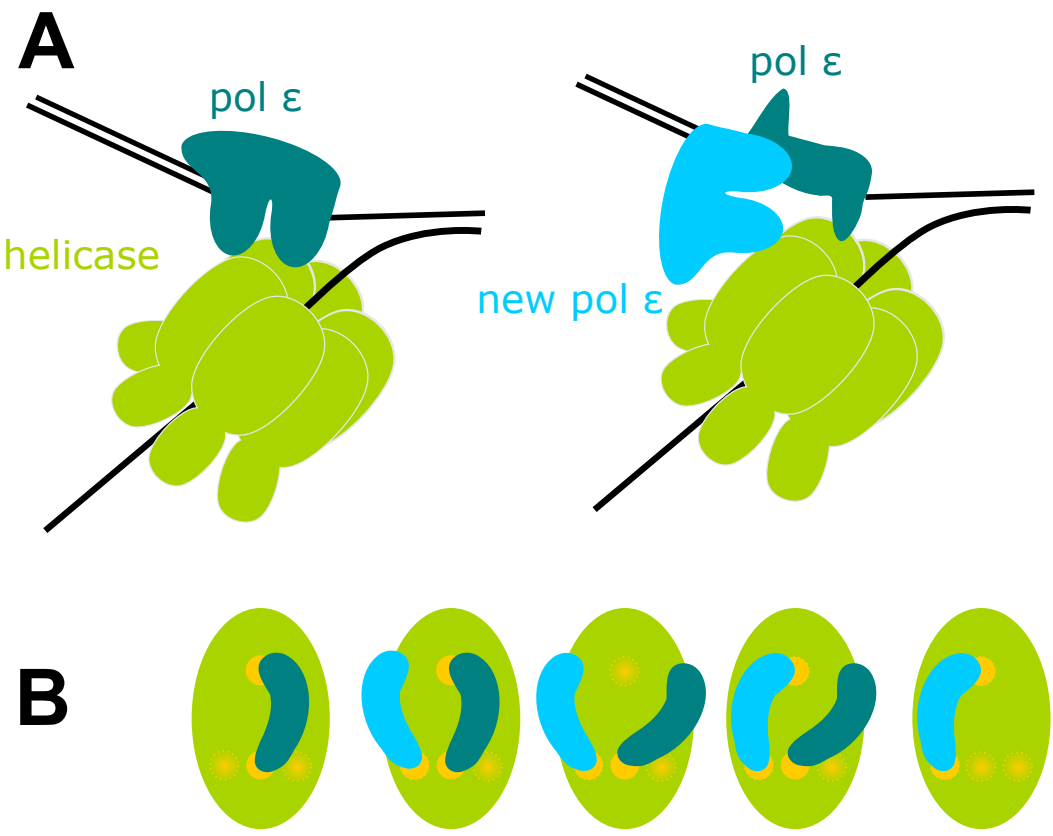



**Figure 26**. **Model for multisite competitive exchange. (A)** Schematic of the binding interaction between DNA polymerase and the CMG helicase. **(B)** The presence of multiple binding sites allows for both a high level of complex stability in the absence of competitor subunits (as a single polymerase subunit can stay tethered through one binding interaction as the other unbinds and rebinds) and rapid subunit exchange in the presence of competitor subunits. Figure adapted from [196].

Aims

The aim of this project is to continue to pursue the question: is the synthesis of the leading strand strictly processive, or can polymerase epsilon subunits be exchanged within a single replicon? To explore the extent to which physiologically relevant polymerase epsilon exchange occurs during ongoing replication in *Saccharomyces cerevisiae*, this project utilizes an *in vivo* inducible system for a functionally analogous but biochemically differentiable Pol2 protein. Next generation sequencing is used as a read out for the replisome incorporation of Pol2 subunits introduced following replication initiation. By utilizing an induction

system which allows for varied expression levels, it is possible to directly assess the *in vitro* findings of tunable levels of polymerase exchange. Finally, given the particular importance of questions regarding the response of the leading strand polymerase to synthesis-blocking DNA lesions, the system can be challenged with DNA damage to assess the impact on polymerase stability at the CMG. This may yield insights into the mechanism of polymerase handoff.

## Results

### Motivation and scope

The aim of this project is to set up a system to assay for switching of polymerase epsilon during ongoing eukaryotic replication *in vivo*. This will build upon *in vitro* work investigating polymerase switching in reconstituted eukaryotic replication [180,188,195] and *in vivo* work that has demonstrated polymerase switching in prokaryotic systems [193,197]. This chapter will begin by laying out the development of an *S. cerevisiae* experimental system to allow for *in vivo* polymerase switching experiments. This chapter will then present cell cycle synchronization ChIP-Seq experiments designed to assay for polymerase switching.

### System development

To assay for polymerase epsilon switching *in vivo*, four major experimental components were required [Figure 27]. First, a rapidly inducible and biochemically differentiable Pol2 protein. Second, a long replicon free of licensed origins to allow for the explicit analysis of polymerase switching events occurring during ongoing replication, rather than showing the incorporation of the induced polymerase into new replication forks. Third, a robust cell cycle synchronization protocol that would work within the experimental system to allow for precise control of *S. cerevisiae* progression through S phase. And finally, the establishment of a method for quantitative ChIP-Seq.

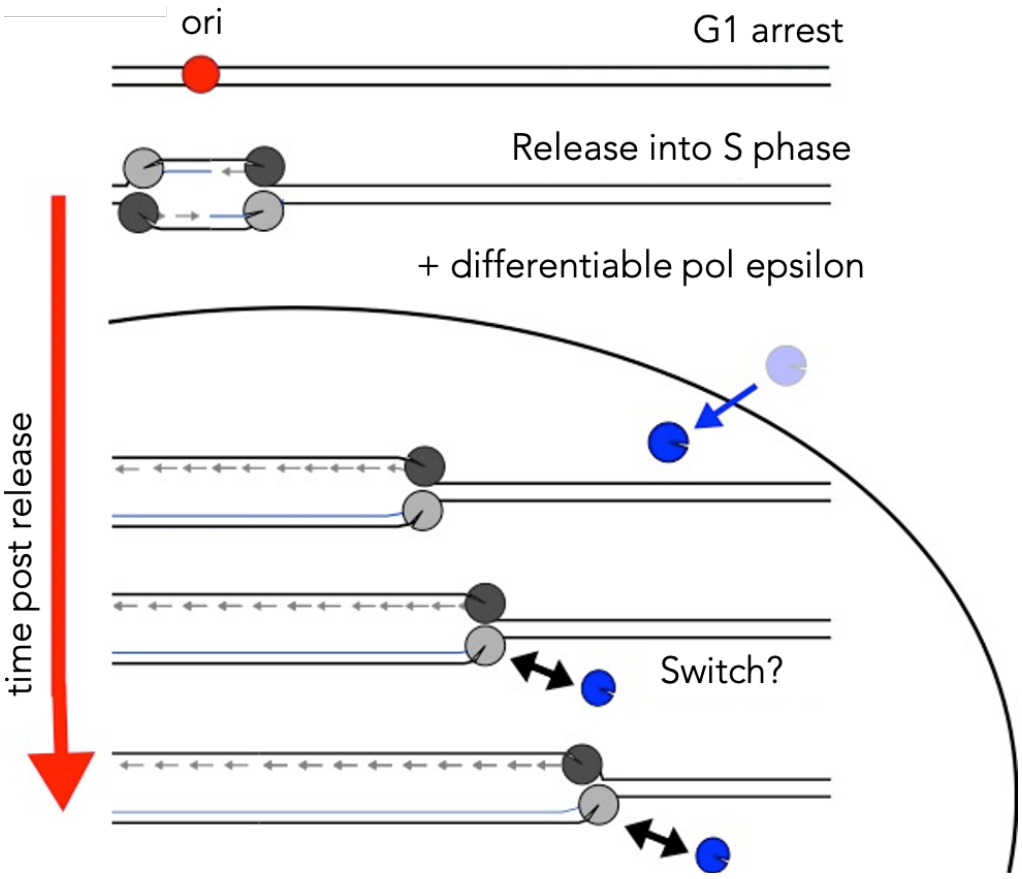



**Figure 27**. **Test for polymerase exchange**. Experimental schematic to assay for polymerase exchange *in vivo*. The system requires (1) an inducible and biochemically differentiable polymerase, (2) a long replicon free of licensed origins of replication, (3) precise control of cell progression into S phase, and (4) a quantitative method for the detection of endogenous versus induced polymerase on the DNA.

### *Inducible Pol2*

#### Localization control

Post-translational control of protein localization was hypothesized to be the most rapid method to allow for the introduction of a new polymerase to the cellular replication machinery after the initiation of replication. Thus, an estradiol localization control approach previously used to induce nuclear localization of a transcription factor in *S. cerevisiae* was adopted for the development of a Pol2 induction system [198,199]. In this method, the hormone binding domain (HBD) of the human estrogen receptor (ER) is added as a tag (termed the 'E tag' here) to a protein of interest. In the absence of estradiol, the HBD is thought to be regulated

by interaction with heat shock proteins (HSP), leaving the E-tagged protein sequestered in the cytoplasm. The addition of beta-estradiol elicits an efficient switch as the HSP interaction is lost, allowing for rapid nuclear localization [Figure 28A]. This system for rapid polymerase induction was introduced into a yeast strain with five origins on the right arm of chromosome VI (*ARS607, 608, 608.5, 609, 610*) inactivated to create an ultra-long replicon (Strain Table: VGY129, kind gift from Donaldson lab). This ultra-long replicon usefully opens up a larger window of the genome – and correspondingly a longer period of time – over which to look for evidence of polymerase switching into ongoing replisomes, as opposed to identifying the incorporation of the newly-induced polymerase into the replisome prior to initiation.

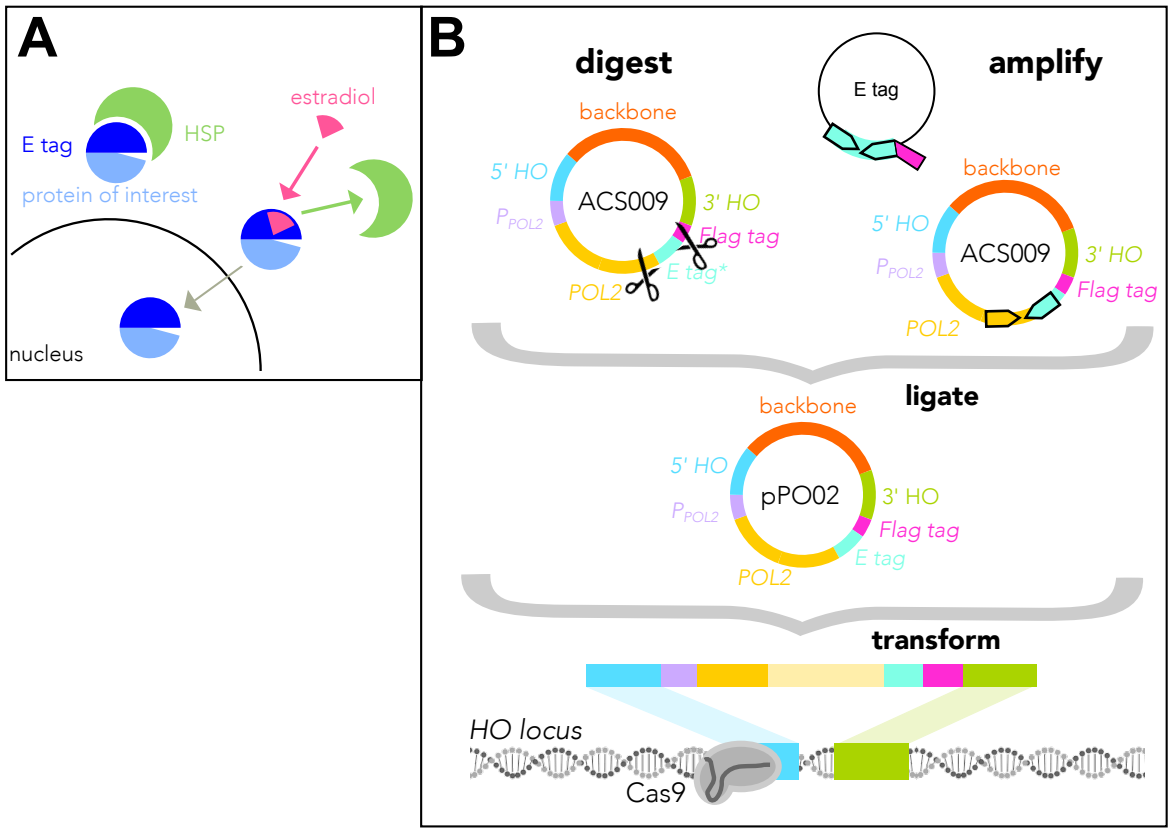



**Figure 28**. **Estradiol localization control inducible Pol2**. **(A)** Estradiol localization control approach. The E tag (estrogen receptor hormone binding domain) fused to the protein of interest drives cytoplasmic sequestration by heat shock protein (HSP) binding. Addition of estradiol releases the binding and allows for nuclear localization and function. **(B)** Schematic for the

creation of a plasmid to introduce a Flag- and E-tagged *POL2* to the *S. cerevisiae HO* locus (Plasmid Table: pPO02). Transformation was carried out with the repair template plasmid pPO02 along with *HO* guide (Plasmid Table: pPO01) and Cas9-Trp plasmid.

To employ this system for Pol2 induction, a Pol2 fusion gene with a C terminal E tag followed by a Flag tag was designed [Figure 28B]. This gene was put under the control of the endogenous *POL2* promoter sequence and flanked on both sides by homology arms mapping to the *S. cerevisiae HO* locus. Ana Wallis, a prior PhD student in the laboratory, carried out the cloning necessary to create a plasmid of this design. However, the E tag in her final product differed from the E tag sequence specified in previous publications. In starting my work on this project, I used the resource that AW created along with a sequence-verified, *de novo* synthesized E tag plasmid as Gibson cloning templates to create the required plasmid. The final product was verified by diagnostic digest (BglII, PacI/SacII), with PCR spanning over junctions (Primer Table: AC0067/CA029, EH248/AC0073, CA1194/EH248, PO49/AC0064), and sequencing (Primer Table: AC0067, EH248), (Plasmid Table: pPO02, pPO03).

Two Cas9 guide RNAs were designed and cloned into the pROS16 plasmid backbone [200] to target double strand breaks to the *HO* locus (Primer Table: AC60, AC61, AC62), (Plasmid Table: pPO01).

The inducible Pol2 repair plasmid, *HO* guide plasmid, and a plasmid encoding Cas9 with Trp selection were introduced through transformation to create strain *POL2-E-Flag*. Transformants were isolated and screened by PCR across the *HO* locus (Primer Table: AC29/CA29, AC29/CA76), (Yeast Table: POY01, POY02). In addition to the introduction of the ectopic inducible *POL2*, the endogenous *POL2* gene was tagged with a 3' Myc tag in both strains with and without the introduced regulatable Pol2 [Figure 29]. This was carried out by amplifying the pYM18

plasmid [201], which encodes the Myc tag followed by a KanMX cassette, with long primers conferring homology to the endogenous *POL2* 3' end (Primer Table: PO47/PO48). This PCR product was transformed into yeast without the requirement for Cas9 or guides as G418 was used to select for cells that had undergone recombination introducing the PCR amplicon. The incorporation at the endogenous locus and the correct genomic sequence at the *HO* locus in the selected transformants were confirmed by PCR (Primer Table: ALU30/PO63, CN127/ALU29, AC29/CA29), resulting in the generation of ultra-long replicon strains with both a Myc-tagged endogenous Pol2 and a Flag-tagged regulatable Pol2 at the *HO* locus (*POL2-E-Flag;POL2-Myc*), and ultra-long replicon strains with the Myc-tagged endogenous Pol2 and a wildtype *HO* locus (*POL2-Myc*) (Yeast Table: POY18, POY19, POY20, POY21).

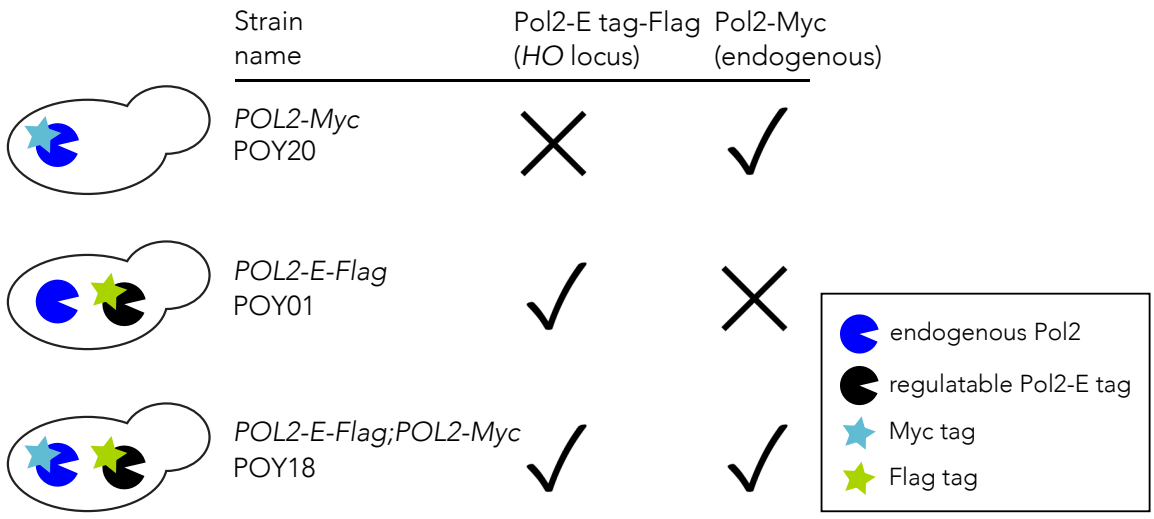



**Figure 29**. **Estradiol localization control Pol2 strain creation**. Creation of strains with endogenous Myc-tagged *POL2*, along with an inducible E- and Flag-tagged *POL2* at the *HO* locus.

The effectiveness of Pol2 localization control was assessed by chromatin fractionation followed by Western blotting for the Pol2-Flag tag. The blot was also probed with an antibody against tubulin to assess the quality of the chromatin

fractionation. In the first experiment, abundant Pol2-Flag was detected in the chromatin fraction in both the presence and absence of estradiol treatment [Figure 30A]. A biological repeat was carried out, and when the Flag band intensity was normalized to the intensity of endogenous Myc-tagged Pol2, it again indicated that the presence of the Flag-tagged Pol2 on chromatin was not dependent upon estradiol treatment [Figure 30B].

These results suggested that the E tag system, as designed, was inadequate for Pol2 localization control. This could be a consequence of the Pol2 protein target. The previously published use of this system in *S. cerevisiae* has been heavily focused on localization control of small and compact transcription factors [199]. Thus, the large and bilobed Pol2 protein or its presence in the polymerase epsilon complex could mask the hormone binding domain tag or be otherwise recalcitrant to cytoplasmic sequestration by heat shock proteins. The Pol2 expression level could also impact the effectiveness of this system: previous work has been restricted to comparatively low-level expression of E-tagged proteins, so it is possible that a higher concentration of the E-tagged protein could saturate sequestration capacity and explain the observed lack of regulation by estradiol.

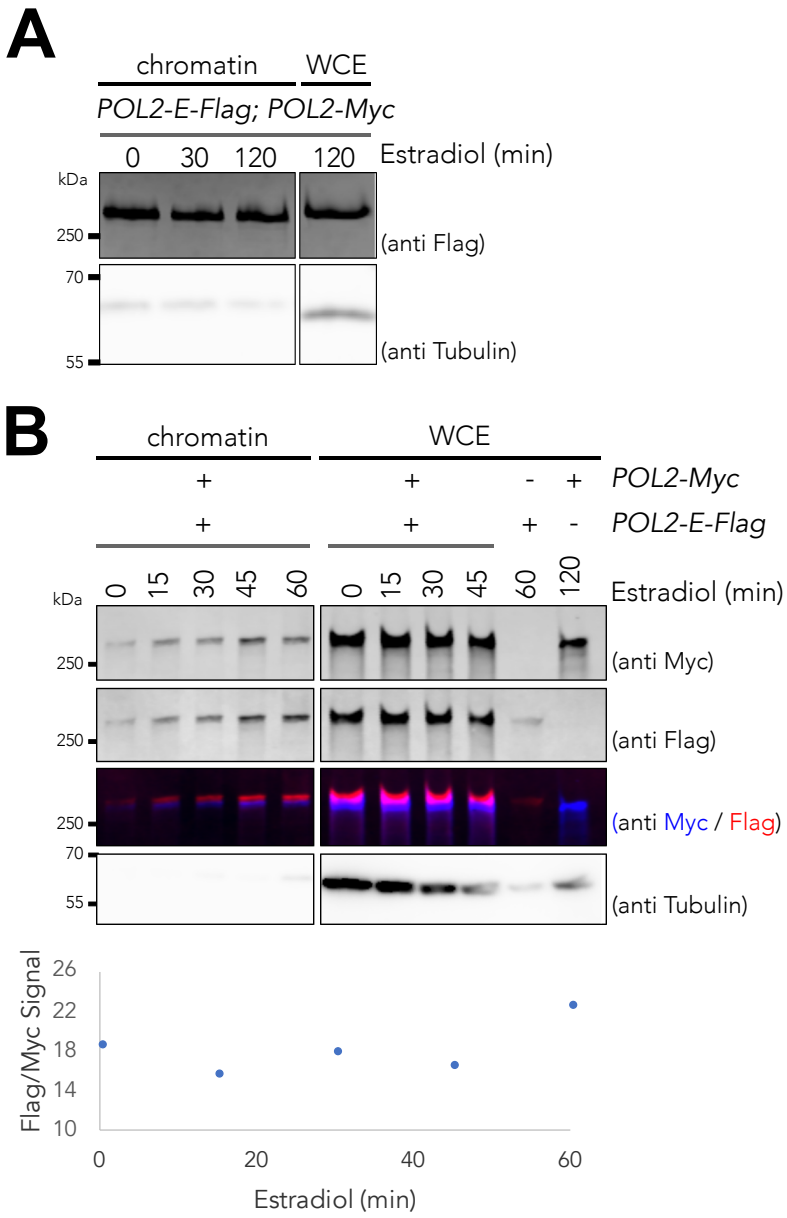



**Figure 30**. **Estradiol localization control effectiveness**. **(A)** *POL2-E-Flag;POL2-Myc* (Strain Table: POY18) was treated with 1 μM estradiol for 0, 30, or 120 minutes. Chromatin and whole cell extract (WCE) fractions were subjected to Western blotting and stained for Flag and tubulin to confirm the effectiveness of the chromatin fractionation. Expected band for Pol2-E-Flag at 287.5 kDa (255.6 kDa Pol2, 34.7 kDa E tag, 6.1 kDa Flag). **(B)** *POL2-E-Flag;POL2-Myc* (POY18), *POL2-Myc* (POY20) and *POL2-E-Flag* (POY01) were treated with 1 μM estradiol for 0, 15, 30, 45, 60, or 120 minutes. Chromatin and whole cell extract (WCE) were subjected to Western blotting and stained for Flag (Pol2-E-Flag), Myc (Pol2-Myc), and tubulin. Flag and Myc band intensities were quantified and the Flag intensity was normalized to the Myc intensity for each chromatin fraction sample to assess the degree of Pol2-E-Flag chromatin localization responsiveness to estradiol treatment.

## Expression control

The second approach tested for achieving regulatable Pol2 was transcriptional activation. In *S. cerevisiae*, galactose-inducible promoters are among the most well-characterized and commonly-used for controlling expression of recombinant genes [202]. Briefly, the *S. cerevisiae GAL* genes including *GAL1*, *GAL7*, and *GAL10* are involved in galactose import and utilization; in the absence of cellular glucose but presence of galactose, genes under *GAL* promoters ($P_{Gal}$) are highly expressed to facilitate cellular metabolism [202]. This expression is repressed by glucose. Long and short (L and S) truncated versions of $P_{Gal1}$ have been characterized as responding to galactose with a lower level of expression activation [203].

To create a system for transcriptional Pol2 induction a plasmid was designed to encode an N-terminally Flag-tagged Pol2 under the control of $P_{Gal1}$, $P_{GalL}$, or $P_{GalS}$ [Figure 31]. This cassette was again flanked by homology arms to allow for *S. cerevisiae* genome integration at the *HO* locus. The MoClo yeast toolkit assembly system was utilized to create the plasmids with each of these Gal promoter variants [20]. Briefly, parts of plasmids encoding the plasmid backbone (pYK089), Pol2 (pEDH16), Flag tag (pYTK060), tTDH1 terminator (pYTK066), and PCR amplicons for the 3' *HO* homology arm (pPO02 amplified with PO86/PO87), 5' *HO* homology arm (pPO02 amplified with PO88/PO89), and the desired *Gal* promoter ($P_{Gal1}$: pYTK030 amplified with PO84/PO85, $P_{GalS}$: pYM-N26 amplified with PO80/PO81, $P_{GalL}$: pYM-N30 amplified with PO82/PO83) were cycled through BsaI digestion and T4 ligation to construct a complete plasmid. The products were

PCR and sequence verified (Primer Table: AC27, PO62, EH287), (Plasmid Table: pPO05, pPO06, pPO07).

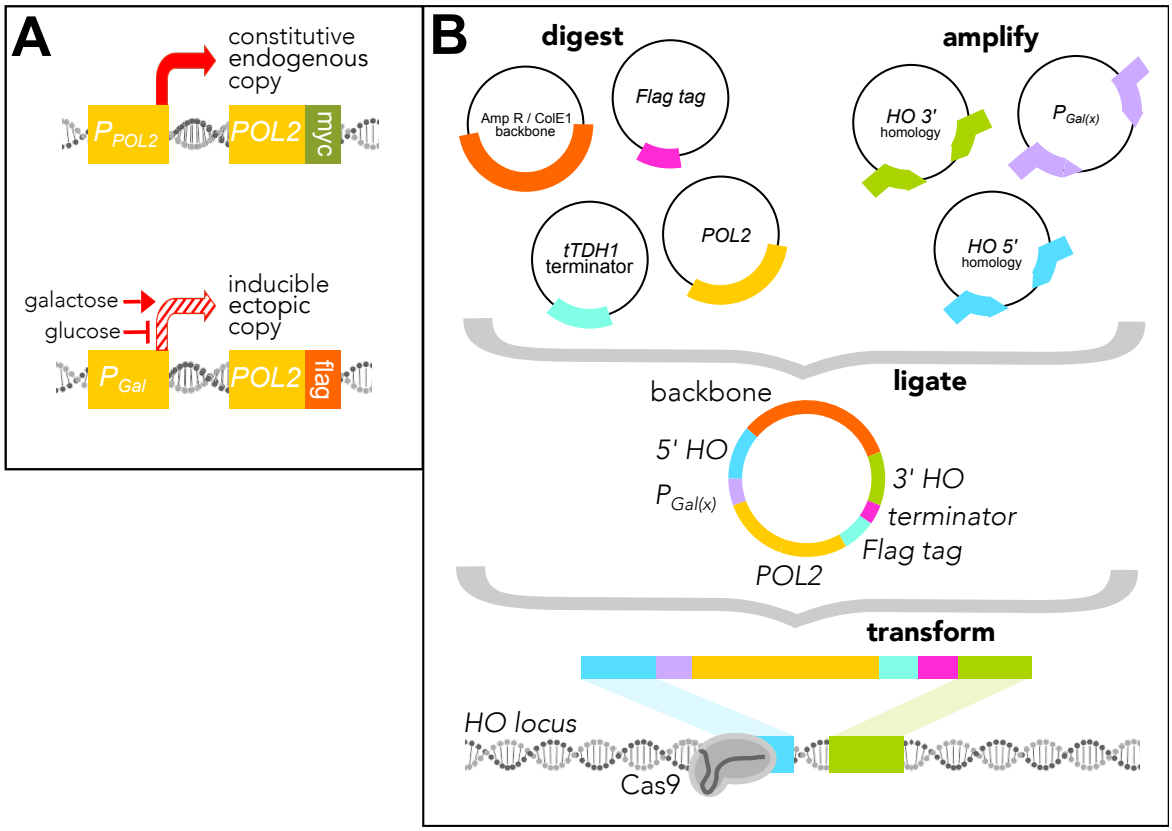



**Figure 31**. **Galactose-inducible expression of Pol2**. **(A)** A Flag-tagged *POL2* under the control of a galactose-inducible promoter ($P_{Gal1}$, $P_{GalS}$, or $P_{GalL}$) is transcriptionally induced by the presence of galactose in the medium and repressed by glucose. The strains additionally encode the endogenous *POL2* gene Myc tagged and remaining under normal cell cycle control coordinated by the *POL2* promoter. **(B)** Schematic for the creation of plasmids to introduce a gene encoding Flag-tagged *POL2* to the *S. cerevisiae HO* locus under a galactose-inducible promoter (Plasmid Table: pPO05, pPO06, pPO07). The transformation was carried out with the repair template plasmid along with an *HO* guide (Plasmid Table: pPO01) and Cas9-Trp plasmid.

These constructs, along with a plasmid encoding Cas9 and the *HO* locus guide plasmid (Plasmid Table: pPO01) were transformed into the *POL2-Myc* ultra-long replicon *S. cerevisiae* strain (Strain Table: POY20) [Figure 32]. The transformants were screened by PCR [Figure 32B] (Primer Table: PO49, AC27) and Western blotting after 8 hours of treatment with 2% galactose to confirm the induction of Flag-tagged Pol2 protein [Figure 32C].

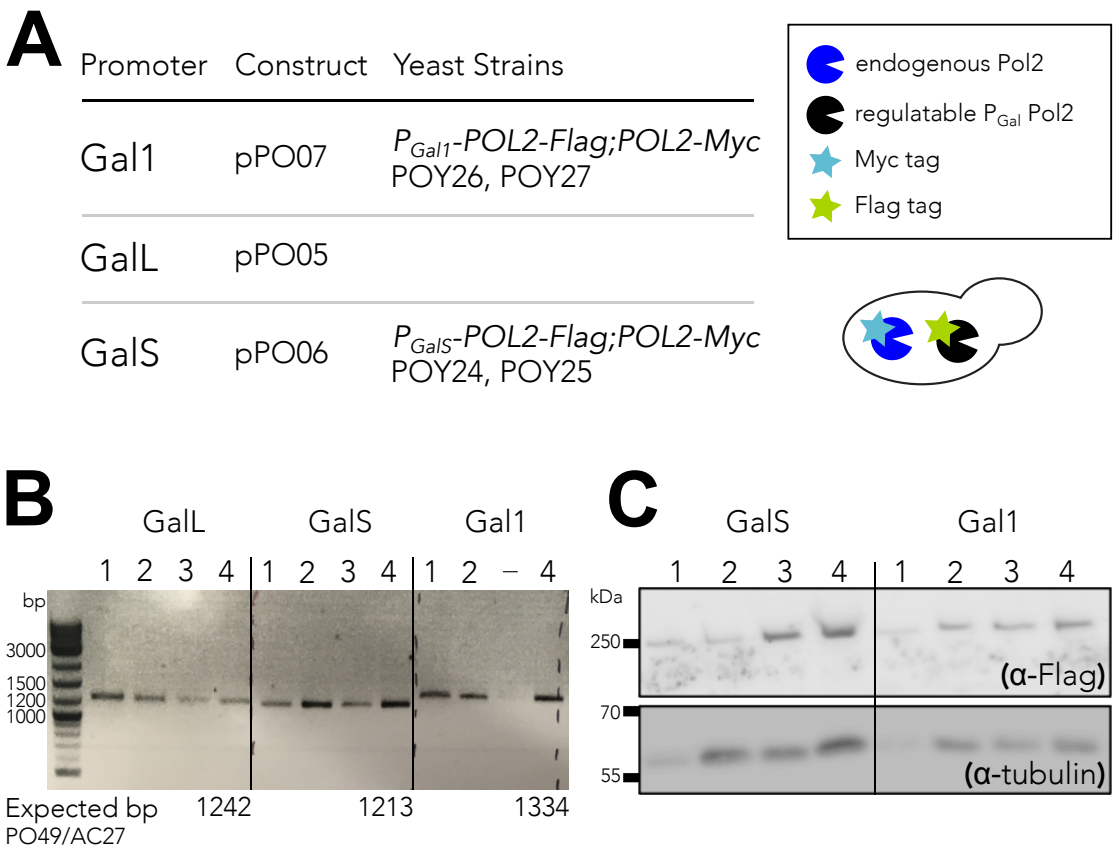



**Figure 32**. **Galactose-inducible expression control Pol2 strain creation**. **(A)** Creation of strains with galactose-inducible expression of Flag-tagged *POL2* at the *HO* locus. All strains additionally had a Myc tag on the endogenous *POL2*. **(B)** Strains were screened for appropriate incorporation of the Flag-tagged *POL2* by PCR from the *HO* locus outside of the plasmid homology arms across to *POL2* (Primer Table: PO49/AC27). **(C)** Further screening for appropriate expression of Flag-tagged Pol2 was carried out via total protein extraction and Western blotting following 8 hours of induction with 2% galactose.

## Gal2 knockout

The galactose promoter is very sensitive to low concentrations of galactose. Knocking out the galactose transporter, *GAL2*, has been demonstrated to allow for tunable control of expression induction in response to varying concentrations of galactose [20], and thus this was engineered into the experimental galactose-induction strain [Figure 33].

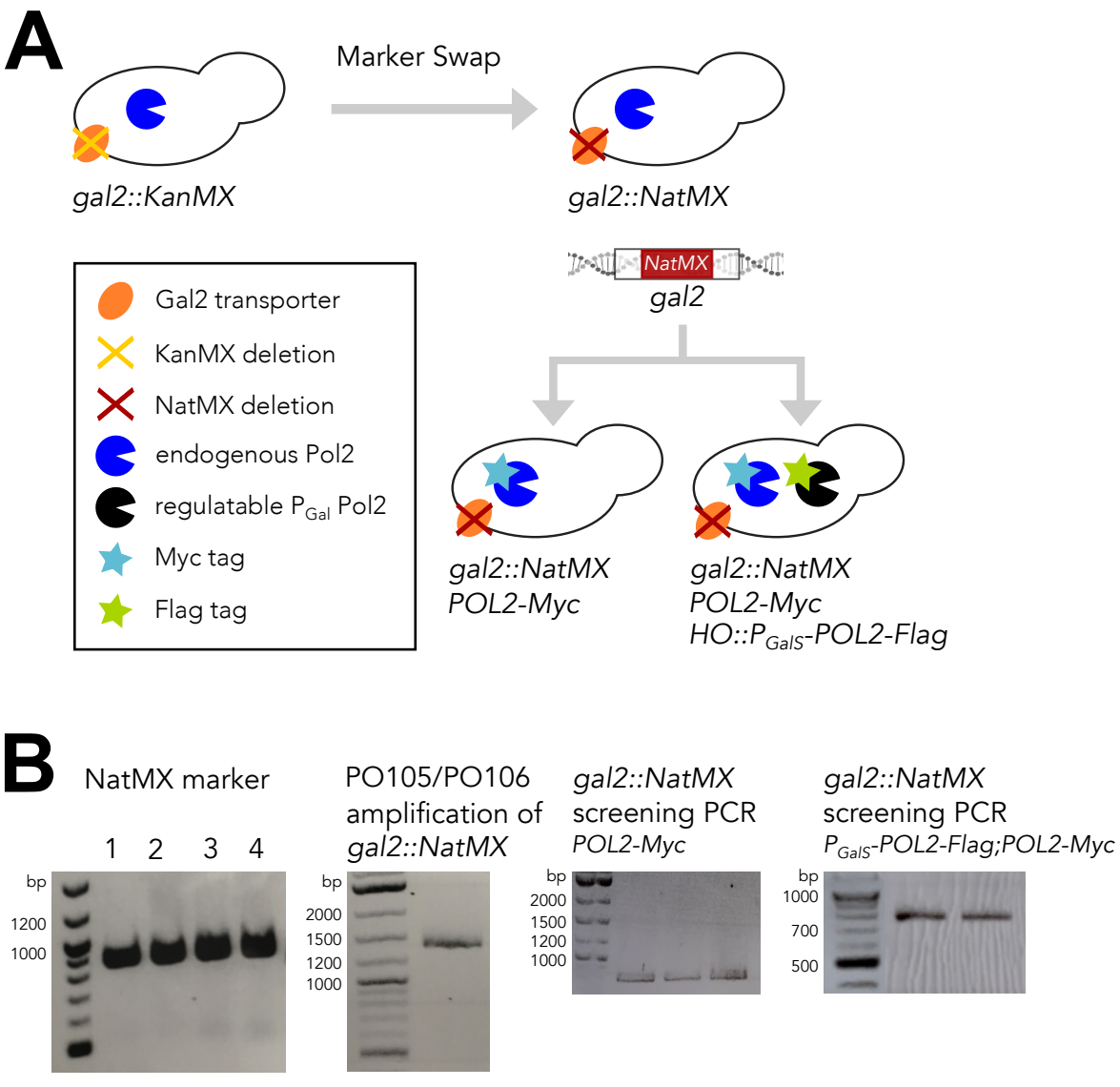



**Figure 33**. ***GAL2* galactose transporter knockout**. **(A)** Strains with the galactose-transporter, *GAL2*, knocked out are expected to show increased sensitivity and tunability for galactose-inducible expression control. To achieve the *gal2* knockout in the experimental strains, a marker swap in a *gal2::KanMX* strain was used to create *gal2::NatMX*. The *gal2* locus was amplified from this new strain and transformed into the experimental *POL2-Myc* and $P_{GalS}$*-POL2-Flag;POL2-Myc* strains. **(B)** Left: marker swap transformants were selected for growth on CloNat, and screened by PCR (Primer Table: PO111, PO112). Middle: *gal2* locus amplification (Primer Table: PO105, PO106). This product was transformed into *POL2-Myc* and $P_{GalS}$*-POL2-Flag;POL2-Myc*. Right: The transformants were selected for growth on CloNat and screened by PCR (Primer Table: PO111, PO112).

The yeast S288C collection has a strain with *GAL2* knocked out via interruption by the *KanMX* cassette. To use this to achieve *gal2* knockout in the experimental strains which already have KanMX resistance at the endogenous *POL2* locus, a marker swap was carried out to replace the *KanMX* cassette conferring G418 resistance with a *NatMX* cassette conferring CloNat resistance. This was

achieved by amplifying the *NatMX* deletion cassette from plasmid pAG25 (Primer Table: CN2169, CN2171). The product of this amplification was transformed into the S288C *gal2::KanMX* strain, with recombination expected as the *KanMX* and *NatMX* deletion cassettes share homology flanking both sides of the resistance marker coding sequence. The correct transformants were selected for with CloNat, screened by PCR (Primer Table: PO111, PO112) [Figure 33B left], and frozen down (Strain Table: POY30, POY31).

The *gal2* locus was amplified from the *gal2::NatMX* strain (Primer Table: PO105, PO106) [Figure 33B center] and the product transformed into *POL2-Myc* and $P_{GalS}$*-POL2-Flag;POL2-Myc* strains. The transformants were selected for with CloNat, and screened by PCR (Primer Table: PO111, PO112) [Figure 33B right]. Successful transformants were frozen down as *POL2-Myc;Δgal2* and $P_{GalS}$*-POL2-Flag;POL2-Myc;Δgal2* (Strain Table: POY32, POY33, POY34, POY35, POY36).

### *Long replicon*

To assay for polymerase switching – rather than looking at the assembly of Pol2 into a new replisome – it was necessary to have the induced polymerase become available on an equivalent timescale to that of the replication fork traversing a single replicon. To ensure that this timing requirement was satisfied, modeling was undertaken to begin to understand 'fork age,' or the length of time that a given replication fork exists from initiation to termination. Two different experimental approaches to push for longer replicons were considered (1) a specific ultra-long replicon on the right arm of chromosome 6 via origin inactivation and (2) genome-wide long replicon creation via time-restricted initiation.

## Fork age modeling

The length of time that a replication fork persists on a chromosome (henceforth its 'age') depends most simply upon the distance it must traverse from its site of initiation to meet a telomere or oncoming fork. As origins do not all fire synchronously, but instead variably fire or do not fire over the course of S phase, the location of the closest oncoming fork to a given point will depend upon when exactly that fork initiated, and thus how much distance it has traversed since its firing time. In the absence of an analytical solution or experimental data reporting explicitly on fork age, a modeling approach was taken to estimate the expected ages of replication forks across the genome.

To do this, the modeling language Beacon Calculus was employed [50]. Briefly, empirically-determined parameters for the positions, efficiencies, and average replication timing of origins on each chromosome were used as input parameters to define probabilities of licensing and initiation for each origin at each point in S phase [51]. 500 simulations were run for each chromosome to capture various possible replication trajectories based on the given probability of each origin's initiation and an assumption of uniform fork progression of 1.4 kilobases / minute. Forks could be tracked over each timestep of a simulation [Figure 34A], and the age of the fork on a particular genomic locus for a given simulation was calculated as the time since that fork initiated. To consolidate these simulations into a metric for each bin in the genome, the median age across all simulations was taken [Figure 34B,C].

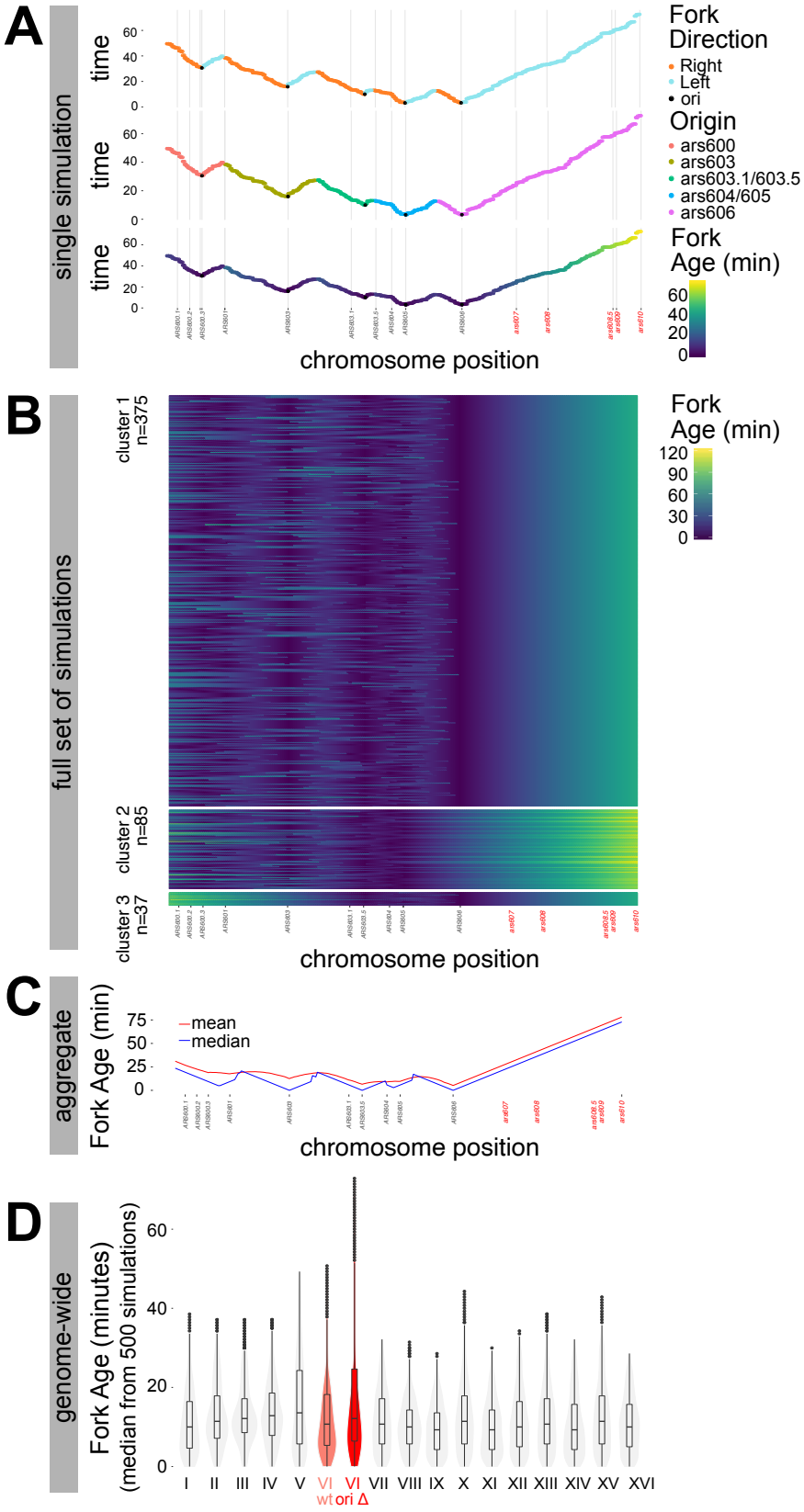



**Figure 34. Beacon calculus was used to simulate replication initiation and progression dynamics**. Parameters for origin location, efficiency, and mean replication timing were taken from [51]. **(A)** Replication was simulated for each *S. cerevisiae* chromosome. From each simulation it was possible to extract replication dynamics such as origin usage, fork direction, and replication timing.

Fork age was calculated by determining the distance between a genomic coordinate and the origin from which its replisome initiated. The simulation shown is an example simulation of chromosome VI with 5 origins inactivated (red). **(B)** 500 replication simulations were run for each chromosome. This heatmap shows the fork ages across chromosome VI with 5 origins inactivated, with simulations clustered by the fork age at each of the origins and across the right arm of chromosome VI. **(C)** An aggregate metric for fork age was generated by taking the median fork age for each genomic bin across the 500 simulations. **(D)** Violin plots of the aggregate (median) fork age densities across each chromosome. The vast majority of the genome was found to be synthesized by forks under 20 minutes old. The 5 origin deletion raised both the 75$^{th}$ percentile and the maximum the replication fork age on chromosome VI.

A caveat to this approach is the assumption of uniform fork speed across the genome. Fork pausing would increase the age of a fork, as it effectively increases the time it takes for a fork to traverse a given distance. The fork speed used in these calculations has been generated from genome-wide data on fork progression, and thus is a reasonable representation of overall fork speed, however specific regions known for difficult fork progression (including heterochromatin) will likely have fork ages underestimated by these calculations.

## Long replicon scenarios

### *Chromosome 6 right*

If the distance that a given fork has to traverse can be increased, the lifetime of that fork on the chromatin will be prolonged, crucially giving more time for a polymerase switching event to occur and a broader genomic range over which to assay for a switch. Given that canonically initiated replication in *S. cerevisiae* has only been characterized as proceeding from active origins, inactivating origins is then an effective way to manipulate the genome to ensure the presence of old forks.

In the work described here, this was achieved by using an *S. cerevisiae* strain with inactivated origins *ars607*, *ars608*, *ars608*.5, *ars609*, and *ars610* on the right arm of chromosome VI. This set of inactivations results in an ultra-long replicon

spanning just over 100 kilobases, from *ARS606* (chrVI: 167,606-168,041) to the telomere (chrVI: 270,161). This yielded an area of particular interest for ChIP-Seq data analysis. Any polymerase reads on the right arm of chromosome VI could be inferred to have come from the fork initiating at *ARS606*, an early firing origin that has been documented to replicate at 18.6 minutes [204], 20.7 minutes [205], or 12.5 minutes into S phase [206].

Modeling supported the hypothesis that deletion of origins on chromosome VI would increase fork age over this region. With the parameters discussed above, modeling showed most regions of the wildtype genome to be replicated by young replication forks, with a median fork age across the genome of 10.7 minutes, and a 75$^{th}$ percentile of 17.1 minutes [Figure 34D]. The maximum fork age in the aggregated simulations of the wildtype genome was 50.7 minutes, which interestingly occurred on the right arm of chromosome VI. This was closely matched by old forks on chromosome V between *ARS517* and *ARS521*, which could be explained by low efficiency and late timing of the intervening origins.

With the inactivation of *ARS607*, *ARS608*, *ARS608.5*, *ARS609*, and *ARS610*, the maximum aggregate replication timing rose to 72.9 minutes as replication of the right arm of chromosome VI came from *ARS606* in 83% of the simulations, and from *ARS605* in those simulations where *ARS606* did not fire [Figure 34B,D].

### *Time restricted initiation*

An orthologous approach to ensuring the presence of old forks throughout S phase is to perturb replication initiation such that only a subset of origins fire, without specifically inactivating origins at the genetic level. This can be achieved with a temperature-sensitive *cdc7* mutant strain: at the permissive temperature

(23°C) replication proceeds unencumbered and this strain can proceed through S phase, however at the non-permissive temperature (38°C) the Cdc7 kinase component of initiator protein complex DDK is inactive, and thus replisome initiation is inhibited [207].

By allowing only a small time-window of Cdc7 activity, it is possible to block replication initiation at all but a subset of characteristically early replication origins. This has been shown to result in slower progression through S phase [207], and is also anticipated to result in longer replicons, as the total number of active forks is reduced and thus the distance that each fork must traverse is increased.

### *Temperature-sensitive cdc7, galactose-inducible POL2-Flag strain creation*

To make use of the 'time restricted initiation' method for ensuring old forks on the chromatin, a yeast crossing strategy was designed to create a galactose-inducible Pol2-Flag, ultra-long replicon strain with the temperature-sensitive *cdc7-1* allele.

Both the temperature-sensitive *cdc7-1* strain (Strain Table: K2032) used in the preliminary double arrest release experiment described below and the experimental *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2* strain were mating type a, and thus a two-generation mating strategy was required to ultimately add the *cdc7* temperature-sensitive allele to the experimental strain [Figure 35A]. The parental *cdc7-1* temperature-sensitive strain was additionally mutant for *bud4* and *rad5*, and thus the spores were screened for wildtype alleles of these two genes.

In the first cross, the *MATa*;*cdc7-1* temperature sensitive strain was crossed with a *MATα* strain with knockouts of *ars607*, *608, 609*, and *610*. The diploid of this cross was sporulated, and spores screened and selected for *cdc7-1* (temperature

sensitivity), origin inactivations (growth on -URA and PCR) (Primer Table: CN2031/CN2036, CN2032/CN2036, DB85/AC39, DB85/AC40, CA53/CA26, CA53/AC42, AC13/AC15, AC14/AC15), and mating type (PCR) (Primer Table: CA378/CA377, CA379/CA377). The spores were also screened for *BUD4* and *RAD5* status (PCR and sequencing) (Primer Table: ALU42/ALU43, ALU17/ALU18).

This first cross and sporulation yielded three candidate haploid progeny with varying combinations of the desired alleles (Strain Table: POY40, POY41, POY42). POY40 was the most promising candidate, as it had the necessary temperature sensitivity and mating type, along with the four origin inactivations and wildtype status for both *BUD4* and *RAD5*. This haploid strain was mated with *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2*.

Work was paused at this point due to the coronavirus.

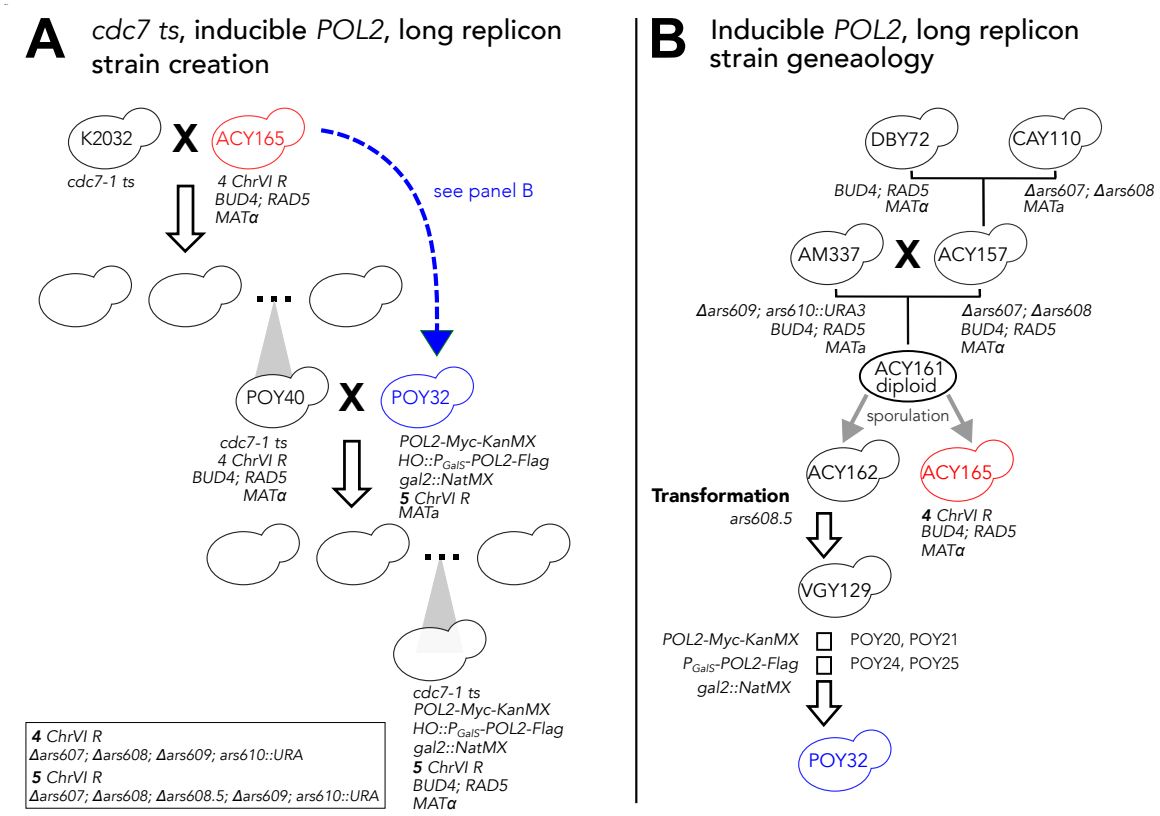



**Figure 35**. **Cdc7 temperature-sensitive experimental strain creation**. **(A)** Mating and screening schematic for creation of the final experimental strain with *cdc7-1* temperature sensitive allele in combination with the inducible Flag-tagged *POL2*, Myc-tagged endogenous *POL2*, *gal2* knockout, ultra-long replicon on the right arm of chromosome VI. **(B)** Genealogy of the *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2* experimental strain, POY32, to indicate relationship to the initial cross strain, ACY165. Strain creation discussed in more detail in the Strains Table.

## *Synchronization*

Tight control over cell cycle progression into S phase is critical to ensure that the induced polymerase becomes available after early origins have already initiated – and thus that induced polymerase presence on the chromatin is indicative of switching as opposed to incorporation at the point of initiation. Additionally, close cell cycle synchrony allows for a better expectation of polymerase locations. In a totally asynchronous cell population, polymerase epsilon is expected to immunoprecipitate across the entire genome, as every base must be replicated at some point over S phase. In a tightly-synchronized S phase cell population, replicative polymerases will begin to act more like a point source chromatin

binding protein, and specific peaks of polymerase localization are expected [Figure 36]. Tighter localization of polymerase epsilon on the chromatin is advantageous for distinguishing true polymerase IP signal from background noise.

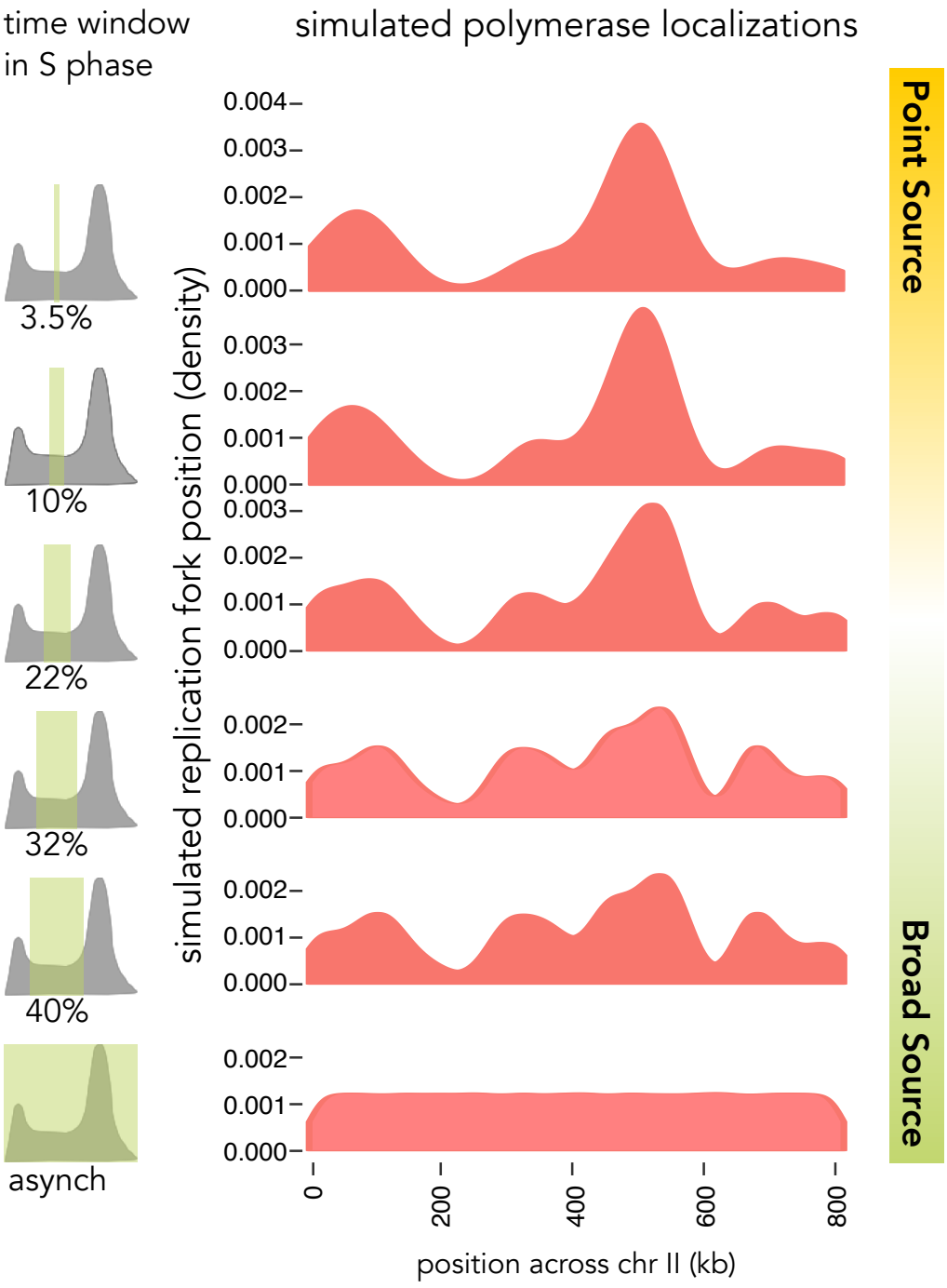



**Figure 36. Beacon calculus simulation of *S. cerevisiae* chromosome II illustrating the behavior of polymerase localization in cell populations representing different S phase synchronies.** 100 simulations of chromosome II were run and polymerase location across the

genome captured for each time step. Timesteps were binned to capture (1) a 5, 15, 30, 45, or 60 minute window (corresponding to 3.5, 10, 22, 32, or 40% of the simulation's replication duration, respectively) centered on the middle of S phase, or all polymerase localizations were captured to reflect an asynchronous population. The polymerase localization for each 'synchronization bin' population was plotted as density of polymerase localization across the chromosome.

To achieve *S. cerevisiae* synchronization, an alpha factor arrest and release protocol was optimized. Given the use of a galactose system for Pol2 transcriptional control, cells had to be grown in a non-preferred sugar source, raffinose, as growth in glucose-containing media opposes galactose induction and would have resulted in a lag in the transcriptional activation. After arrest in G1 with alpha factor, cells were released with pronase or by filtration to degrade or remove the alpha factor, respectively [Figure 37]. Treatment with 200µg/mL pronase or filtration were both sufficient to allow for release and progression through S phase, with DNA synthesis becoming evident 30 minutes after pronase release in both cases [Figure 37B,C]. While cells were able to release and progress through S phase, the synchrony was not sufficiently tight to allow for high-resolution localization of polymerases in ongoing replication. The growth temperature was dropped to 23°C to slow cell cycle progression, and under these conditions the synchrony was marginally improved [Figure 37D]. This problem of poor alpha factor release synchrony has been previously observed for *S. cerevisiae* grown in raffinose (E. Heron, personal correspondence), and thus it was determined that additional synchronization controls were necessary.

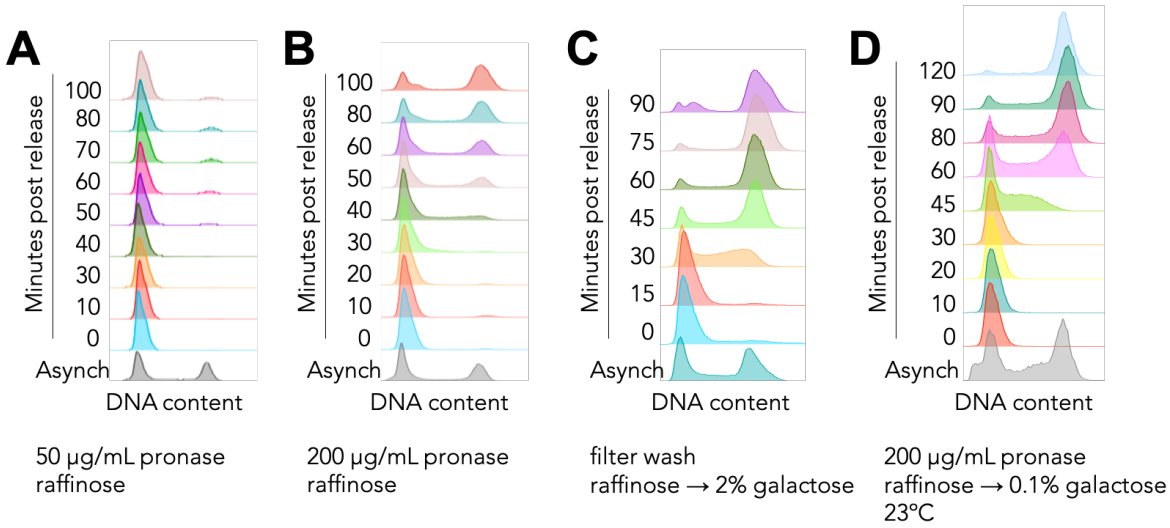



**Figure 37. Alpha factor cell cycle synchronization optimization. (A-C)** Flow cytometry profiles of DNA content of fixed cells as measured by SYTOX staining. Cells were grown asynchronously in raffinose at 30°C at the start of the experiments and were arrested with alpha factor and monitored by microscopy for successful arrest. Cells were released with: **(A)** 50 µg/mL pronase, **(B)** 200 µg/mL pronase, **(C)** filter-washing to remove the alpha factor, followed by re-suspension in 2% galactose to assess synchrony under the conditions needed for galactose induction experiments. **(D)** Cells asynchronously grown in raffinose at 23°C were arrested with alpha factor and monitored by microscopy for successful arrest. Cells were released with 200µg/mL pronase, and the medium supplemented with 0.1% galactose.

It was hypothesized that in addition to offering the opportunity for long replicons genome-wide, the temperature-sensitive *cdc7* mutant might also allow for improved synchrony through utilization of a double arrest release method. This potential for improved synchrony was tested by taking the *cdc7-ts S. cerevisiae* strain, K2032, through four block-release protocols [Figure 38]. First, cells were grown asynchronously in raffinose at the permissive temperature. Alpha factor was added to arrest cells in G1 at the permissive 23°C, and when the arrest was achieved the culture was split into four different cell cycle control regimens.

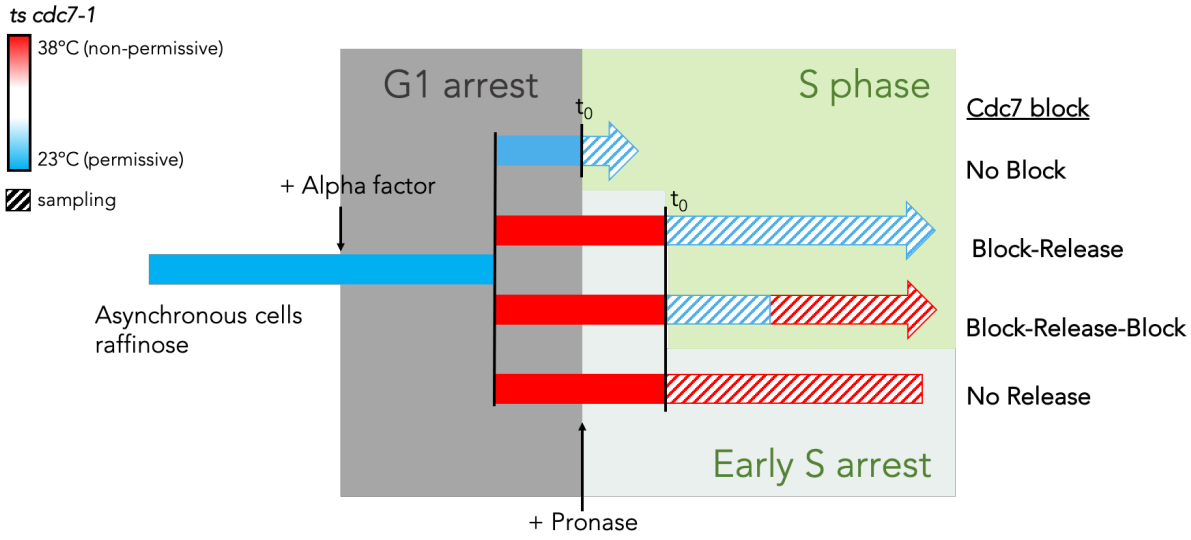



**Figure 38. Schematic for double arrest-release**. A temperature sensitive *cdc7* mutant *S. cerevisiae* strain (Strain Table: K2032) in which replication initiation is inhibited by non-functional Cdc7 kinase at the non-permissive temperature (38°C) was utilized. An experiment was planned with four arrest-release protocols to assess the efficacy of a double arrest release for achieving synchrony in progression through S phase for cells in raffinose.

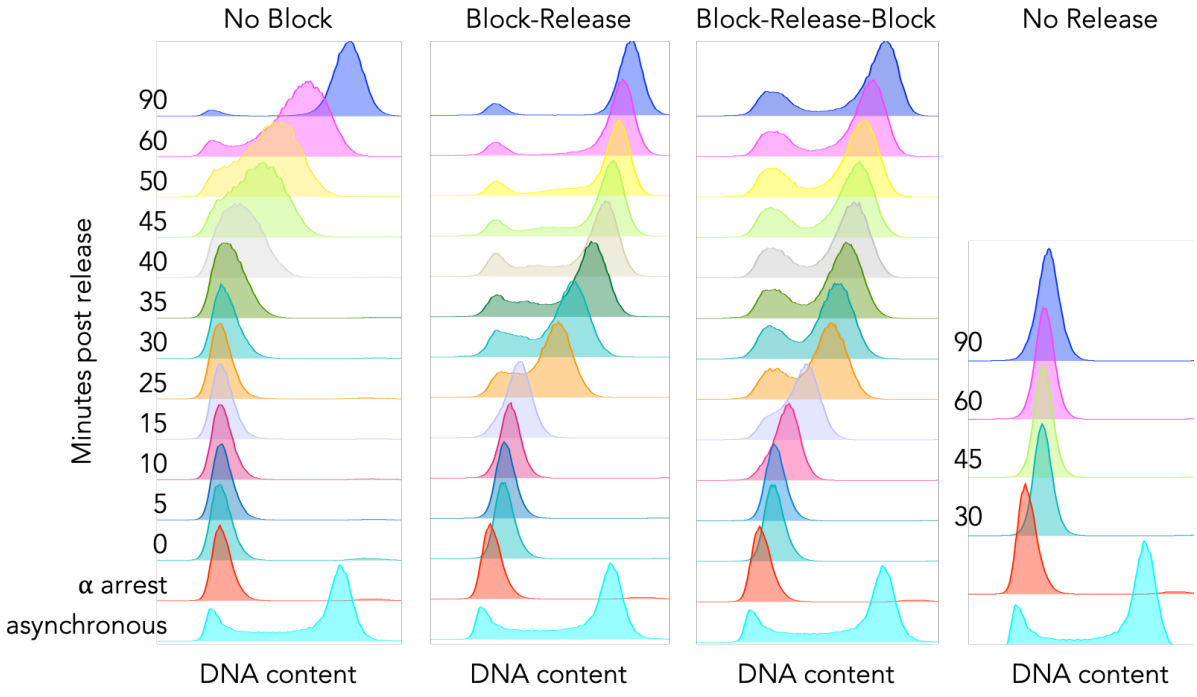



**Figure 39. Double arrest-release synchronization kinetics**. Flow cytometry analysis of cell cycle progression following alpha factor, Cdc7 double arrest-release protocol outlined in [Figure 38]. DNA content assessed by SYTOX staining.

One aliquot was left at 23°C and pronase (200 µg/mL) was added ('No Block'), with $t_0$ defined as the point of pronase addition. The next three aliquots were all raised to the non-permissive temperature to inhibit Cdc7. After incubation at the

non-permissive temperature, pronase (200 µg/mL) was added to release cells from the alpha factor arrest while holding them in the Cdc7 arrest. The cultures were kept at the non-permissive temperature and monitored for successful progression into the start of S phase absent of DNA replication (>95% of the culture large-budded). At this point ($t_0$), one of the three aliquots was dropped to the permissive temperature and kept at the permissive temperature for the remainder of the experiment ("Block-Release") while samples were taken for flow cytometry. The second aliquot was dropped to the permissive temperature for 5 minutes, and then returned to 38°C ("Block-Release-Block") for the remainder of the experiment. The final aliquot was never dropped from the non-permissive temperature ("No Release") and samples were taken at intervals for flow cytometry to assess the degree of leakiness of the Cdc7 inhibition at 38°C.

The 'No Block' [Figure 39] population closely resembled those in previous alpha factor arrest-release experiments, with DNA synthesis becoming apparent by 35 minutes following pronase release. 'No Block' cells progressed as a broad peak to G2 by 90 minutes post release. The addition of a Cdc7 block and release ('Block-Release') improved the synchrony markedly [Figure 39]. The restriction of Cdc7 activity to a 5 minute window (Block-Release-Block) – a protocol which could be used for the analysis of long replicons – similarly resulted in tight synchrony though S phase. This treatment also resulted in a larger residual peak at G1, as any cells that had not initiated replication in the 5 minute window of permissive temperature were blocked from late entry into replication.

The cells of the 'No Release' arm remained at a constant peak location throughout the entire period of sampling post release from alpha factor arrest,

indicating the effectiveness of the temperature sensitive Cdc7 in blocking replication initiation. Of note, the small shift seen in the DNA content peak up from the position of the G1 peak is likely indicative of mitochondrial DNA replication upon release from alpha factor arrest [208].

### *Calibrated ChIP-Seq protocol*

ChIP-Seq is a powerful method for mapping a protein's localization on a genome, but in the absence of a normalization technique it is not effective for measuring changes in total protein abundance on the chromatin. As the experimental question depends upon discerning a change in the prevalence of a target population (the induced Flag-tagged Pol2), it is critical to use an assay that allows for an accurate relative quantitation of the target population. Thus, a calibrated ChIP-Seq protocol was designed in which a genetically distinct yeast species was spiked-in to the experimental sample [Figure 40]. The spiked-in chromatin could then be subjected to immunoprecipitation, and the number of reads mapping to the spike-in genome used as a calibration factor for the reads mapping to the experimental genome [43].

This approach required a yeast species of sufficient similarity to *S. cerevisiae* that the same experimental conditions for cell wall digestion and chromatin extraction would be effective, but sufficiently genetically distinct to allow for sequencing reads to be unambiguously mapped to either the spike-in or the experimental genome. For this, *Saccharomyces bayanus* was chosen.

Furthermore, for this approach to be effective, the immunoprecipitation must pull down chromatin-bound protein from both the experimental and the spike-in strain. Previous work in the lab had been done by Dr. Carolin Müller to create an *S.*

*bayanus* strain with a C terminal Flag tag on Mcm4 (CAY300). Thus, this strain was an ideal spike-in for the Flag IP for induced Pol2 in *S. cerevisiae*. To create a spike-in control for both the Flag (induced Pol2) and the Myc (endogenous Pol2) immunoprecipitations, a double-tagged *S. bayanus* strain was required. A plasmid was designed which would allow for 3' addition of Flag and Myc tags, selected for with kanamycin resistance [Figure 41A]. This plasmid was created through a Gibson reaction and verified through diagnostic digest and PCR (Primer Table: EH251, CN2198). The final plasmid (Plasmid Table: pPO08, pPO09) was amplified to confer homology arms to the *S. bayanus MCM4* 3' end (Primer Table: CA596, CA597), and the product transformed into wildtype *S. bayanus*. Correct transformants (Strain Table: POY37, POY38, POY39) were selected with G418, screened for with primers spanning the *MCM4* 3' end [Figure 41B] (Primer Table: CA605, CA606), and confirmed via Western blotting [Figure 41C].

Several ChIP-Seq experiments were carried out while the double spike-in strain creation was underway. In these experiments the singly tagged *S. bayanus MCM4-Flag* strain (CAY300) was used as a spike-in for induced Pol2 read calibration. As immunoprecipitation unavoidably carries through some background level of nonspecific binding, the Myc immunoprecipitation (IP) sequencing yielded a small fraction of reads that mapped uniquely to the *S. bayanus* genome, despite the absence of a Myc-tagged *S. bayanus* protein. Under the rationale that the level of background binding for the *S. bayanus* chromatin would be equivalent across samples given that they were processed simultaneously through the same IP protocol, this background read number was used as a proxy for normalization of the *S. cerevisiae* reads in the Myc-IP samples.

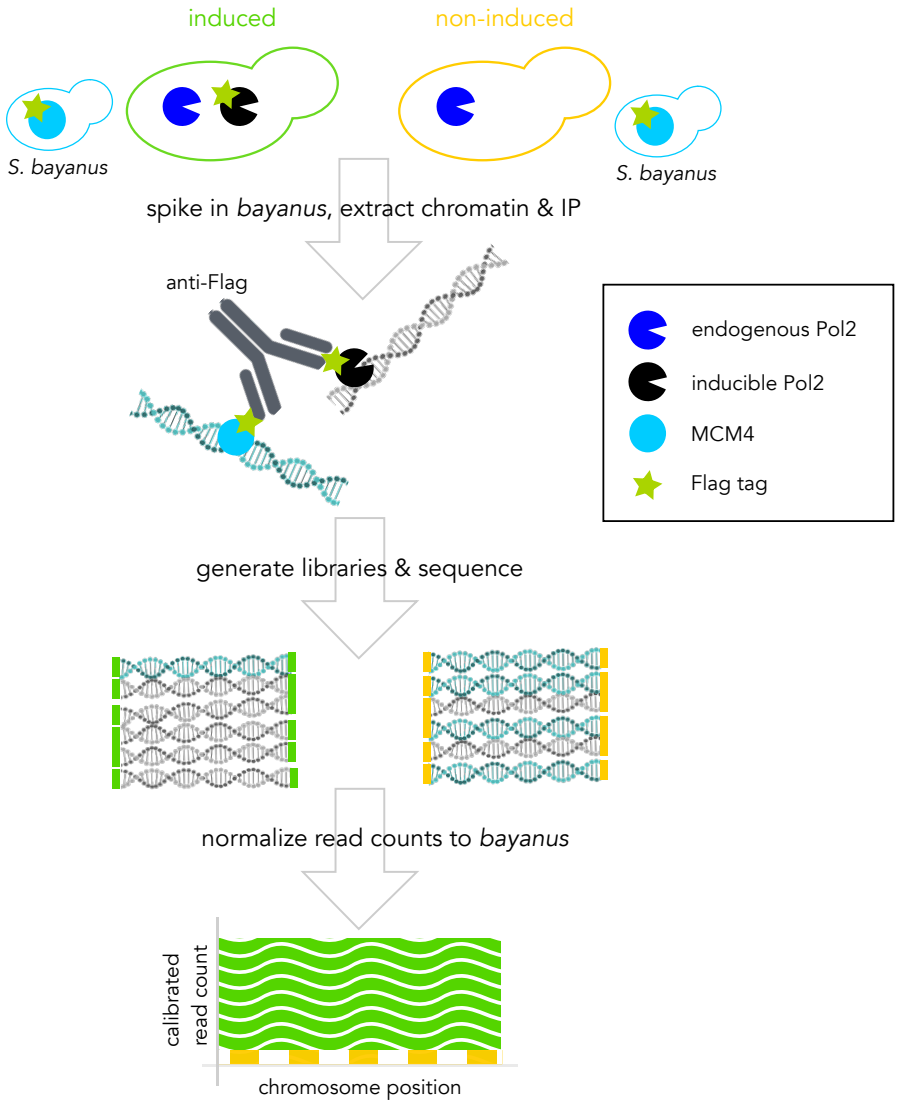



**Figure 40**. **Calibrated ChIP-Seq approach**. To allow for accurate assessment of protein abundance a spike-in method was utilized wherein a closely related yeast, *S. bayanus*, was genome-edited to add a C terminal Flag tag to the Mcm4 protein. The *MCM4-Flag S. bayanus* culture is spiked into the *S. cerevisiae* samples at a cellular ratio of 5:95 *bayanus:cerevisiae* before chromatin extraction and immunoprecipitation. The sequenced reads for each sample can be mapped uniquely to either the spike-in or experimental genome, and the ratio of *S. cerevisiae* to *S. bayanus* reads can inform about the original abundance of the tagged protein in the experimental sample. For a sample in which chromatin occupancy of the Flag-tagged protein in the *S. cerevisiae* cells closely matches the occupancy of the Flag-tagged protein in the *S. bayanus* cells, the ratio is expected to be near to the original cellular ratio of 5:95. In an *S. cerevisiae* sample that lacks any Flag-tagged protein on the chromatin, the *S. bayanus* reads are expected to predominate, as the immunoprecipitation will pull down more of the Flag-tagged *S. bayanus* protein than Flag-tagged *S. cerevisiae* protein.

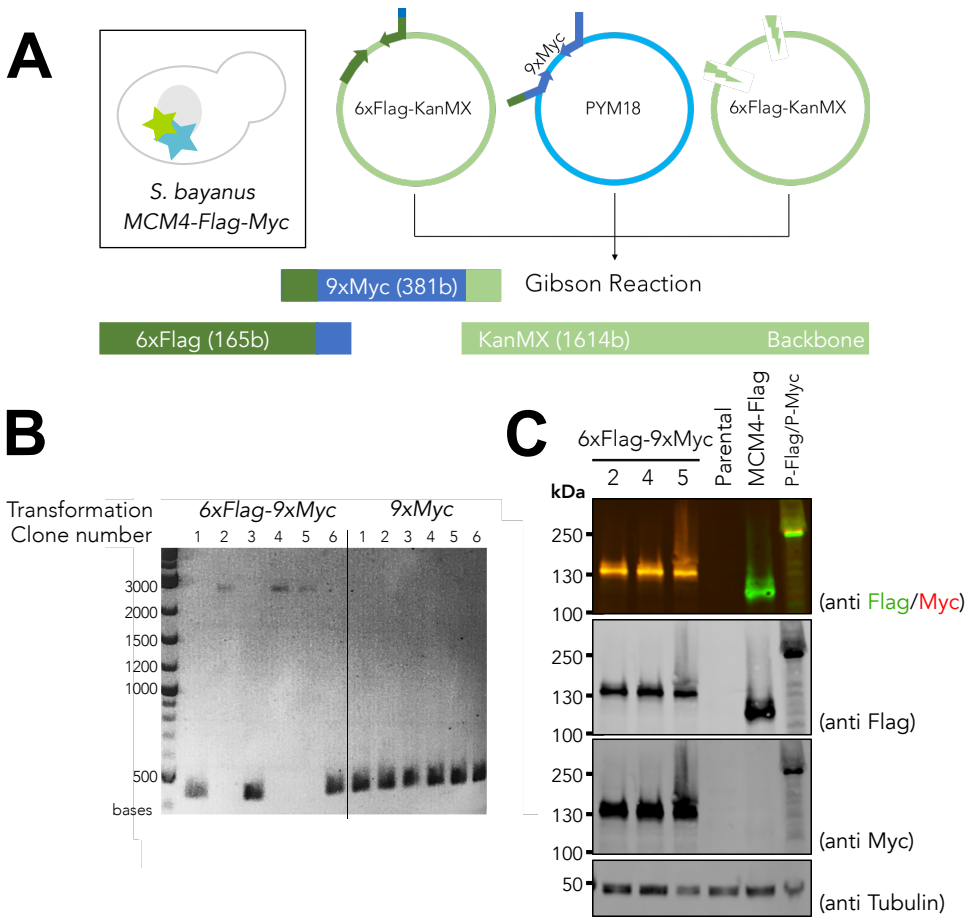



**Figure 41**. **Dual spike-in strain creation**. **(A)** Gibson cloning was used to create a plasmid (Plasmid Table: pPO08, pPO09) encoding a *6xFlag-9xMyc* tag followed by the *KanMX* cassette. This construct was amplified with primers conferring homology arms for *S. bayanus MCM4*, allowing for the creation of a double tagged spike-in strain effective for calibrating read counts for both Flag and Myc IPs. *S. bayanus* strain JRY8149 was incubated with the transforming DNA at 37°C for 30 minutes to achieve transformation, and selection was carried out by replica plating onto G418. **(B)** JRY8149 transformants of either an amplicon of *6xFlag-9xMyc* construct or a *9xMyc* construct were screened with PCR spanning across the *MCM4* 3' end (Primer Table: CA605, CA606). Expected bands: wildtype – 442 base pairs, *MCM4-Flag-Myc* – 2549 base pairs, *MCM4-Myc* – 2266 base pairs. The clones that showed the correct product (*i.e.* clones 2, 4, and 5) were frozen down (Strain Table: POY37, POY38, POY39). **(C)** Western of total protein extraction, probed for Myc and Flag to confirm effective double-tagging of the Mcm4 protein. Samples include clones 2, 4, and 5 from the *5xFlag-9Myc* transformed *S. bayanus* strain (Strain Table: POY37, POY38, POY39), the parental *S. bayanus* strain (JRY8149), the single spike-in *MCM4-Flag* strain (CAY300), and a *POL2-E-Flag;POL2-Myc* sample as a positive control for the antibody staining.

## System characterization

### *Speed of induction*

To identify polymerase switching over the course of ongoing replication, the induction needs to be achieved on a timescale commensurate with replication across a single replicon. The initial characterization of the galactose-responsive

Flag-tagged Pol2 strain showed high levels of induction following 8 hours at 2% galactose [Figure 32C], but further work was required to more precisely understand the induction kinetics.

Two independent isolates of the *$P_{Gal1}$-POL2-Flag;POL2-Myc* (Strain Table: POY26, POY27) and the *$P_{GalS}$-POL2-Flag;POL2-Myc* (Strain Table: POY24, POY25) strains were assessed for their response to induction upon a change from growth in raffinose to 2% galactose (*N.B.* this was performed before the *gal2* knockout was completed in any strains). Each of the strains showed an absence of Flag-tagged Pol2 in growth in raffinose, as assayed by Western blotting of total protein extract [Figure 42]. Furthermore, as previously seen, by 2 hours in galactose, each of the strains showed a robust induction of the Flag-tagged Pol2 [Figure 42]. By 15 minutes in galactose, the two strains with Pol2-Flag expressed from the GalS promoter already showed a Flag-tagged Pol2 band [Figure 42]. By 30 minutes, Pol2-Flag induction was achieved using both the GalS and Gal1 promoters [Figure 42].

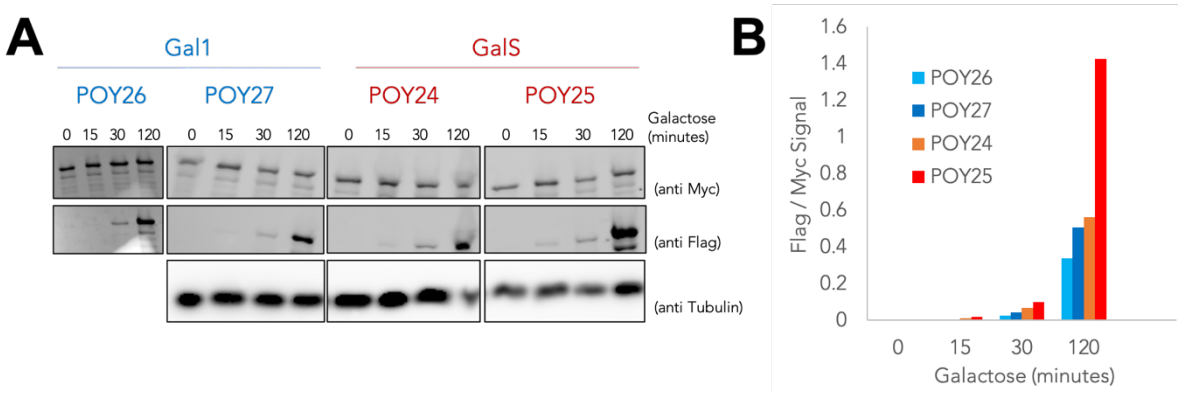



**Figure 42. The kinetics of galactose-induced Pol2 expression from *Gal1* and *GalS* promoters in asynchronously growing cells.** (A) *$P_{Gal1}$-POL2-Flag;POL2-Myc* (Strain Table: POY26, POY27) and *$P_{GalS}$-POL2-Flag;POL2-Myc* cells (Strain Table: POY24, POY25) were grown asynchronously in raffinose, and Pol2-Flag expression induced with the addition of galactose to a final concentration of 2% for the times indicated. Total protein was extracted and subjected to Western blotting. (B) Pol2-Flag signal was quantified using LiCOR software and normalized to the endogenous Pol2-Myc signal as a loading control.

The *$P_{GalS}$-POL2-Flag;POL2-Myc* strain with the highest expression induction (Strain Table: POY25) was taken forward for further analysis. A repeat of the induction with 2% galactose following asynchronous growth in raffinose again showed induction by 20 minutes, with further increase in signal up to 2 hours [Figure 43].

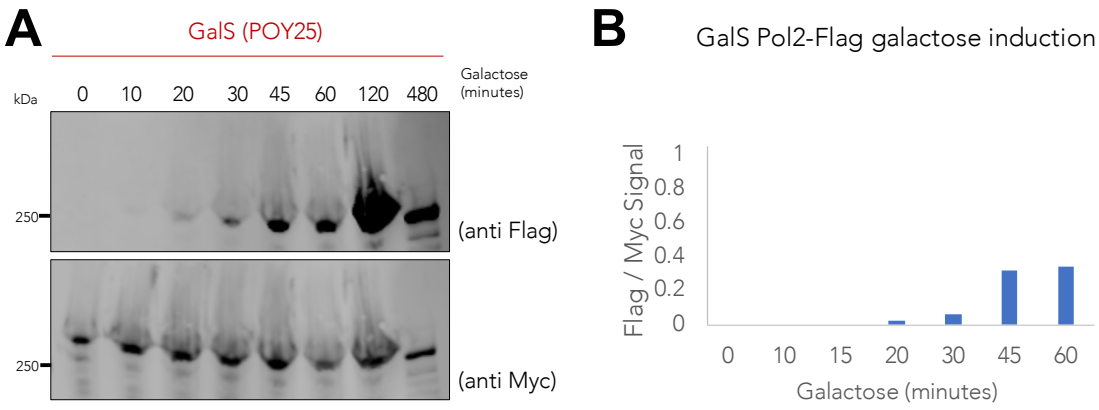



**Figure 43. The kinetics of galactose-induced *POL2* expression from the *GalS* promoter in asynchronously growing cells. (A)** *$P_{GalS}$-POL2-Flag;POL2-Myc* cells (Strain Table: POY25) were grown asynchronously in raffinose, and then Pol2-Flag expression induced with the addition of galactose to a final concentration of 2% for the times indicated. Total protein was extracted and run on a Western blot stained for inducible Pol2 (Flag) and endogenous Pol2 (Myc). **(B)** The endogenous Pol2-Myc and induced Pol2-Flag signals were quantified using LiCOR software, and the endogenous Pol2-Myc signal was used as a loading control for normalization of induced Pol2-Flag levels. The 120 minute sample was not quantified due to overloading.

As the induction ultimately needed to be carried out not in asynchronously-dividing cells, but rather in cells upon release into S phase, the expression kinetics were assayed in cells induced with galactose simultaneously with release from alpha factor arrest [Figure 44A]. Similarly to previous alpha factor-based synchronizations in raffinose at 30ºC, cells showed evidence of DNA replication by 30 minutes following release by addition of pronase [Figure 44B]. Furthermore, by 30 minutes following release, inducible Pol2-Flag could be detected by Western blotting [Figure 44C].

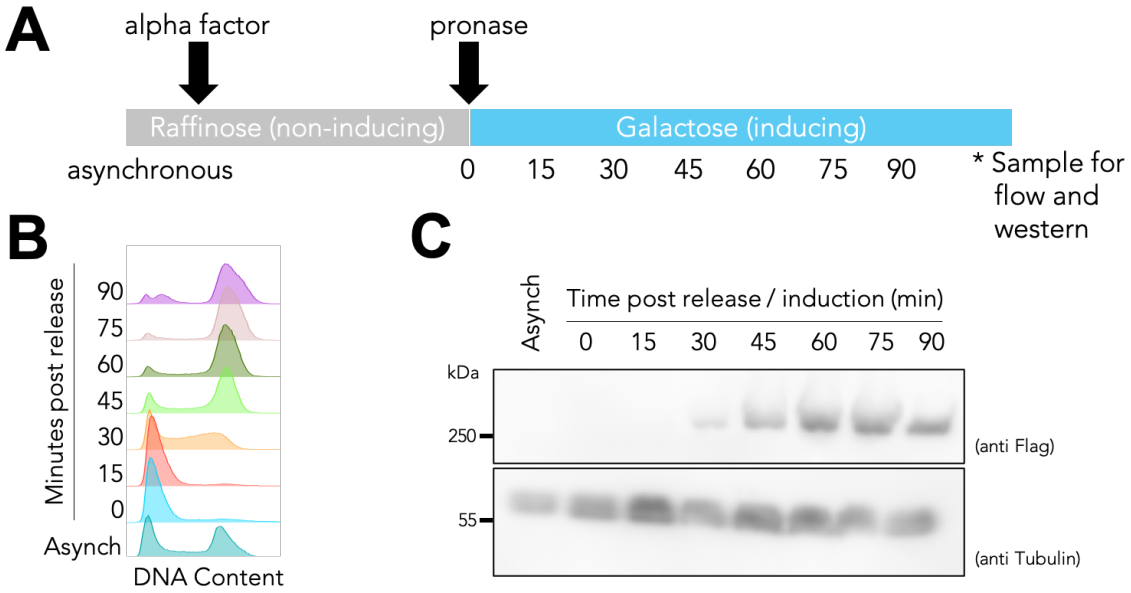



**Figure 44. Galactose-induced *POL2* expression kinetics upon S phase entry. (A)** *$P_{GalS}$-POL2-Flag;POL2-Myc* (POY25) was grown asynchronously in raffinose and subjected to an alpha factor arrest and induced with galactose at the point of release into S phase. Samples were taken for flow cytometry and Western blotting from the asynchronous culture, the arrested culture, and in 15 minute intervals following pronase release into S phase. **(B)** Flow cytometry analysis of DNA content for the arrest and release into S phase. **(C)** Western blot of total protein samples from cells throughout the arrest-release, probed for inducible Pol2 (Flag) and tubulin as a loading control.

### *Quantification of induction level*

To fully understand the potential for Pol2 switching and the impact of the Pol2 induction, it was important to quantify the level of Pol2-Flag expression relative to endogenous Pol2 levels following galactose treatment. To do so, a normalization control strain, Flag Norm, was created to put a Flag tag on the endogenous Pol2 (Strain Table: POY28, POY29). Total protein extract from this strain grown asynchronously could then be loaded as a normalization sample in Western blots with the galactose-inducible Pol2 samples. This allows for a direct quantitative comparison using the same antibody between the level of ectopic galactose-induced Pol2-Flag in the experimental strain and to the level of endogenous Pol2-Flag in the normalization strain.

This experiment was carried out with two different clones of the Flag Norm strain versus *$P_{GalS}$-POL2-Flag;POL2-Myc* and *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2* grown asynchronously in raffinose and then induced with 2% galactose for 2 hours [Figure 45]. The tubulin signal for each sample was used as a loading control for Flag signal and then the tubulin-normalized Flag signal for experimental strains was compared to the same from the normalization Flag Norm strains. Two technical repeats showed the expression of the galactose-induced Pol2-Flag to be between 2 and 6-fold higher than the expression of the endogenous Pol2. While the galactose transporter knockout is anticipated to tune galactose induction (and thus result in lower Pol2 levels in the *Δgal2* knockout strain versus the *gal2* wildtype strain), this effect is not anticipated at galactose levels as high as 2% [20].

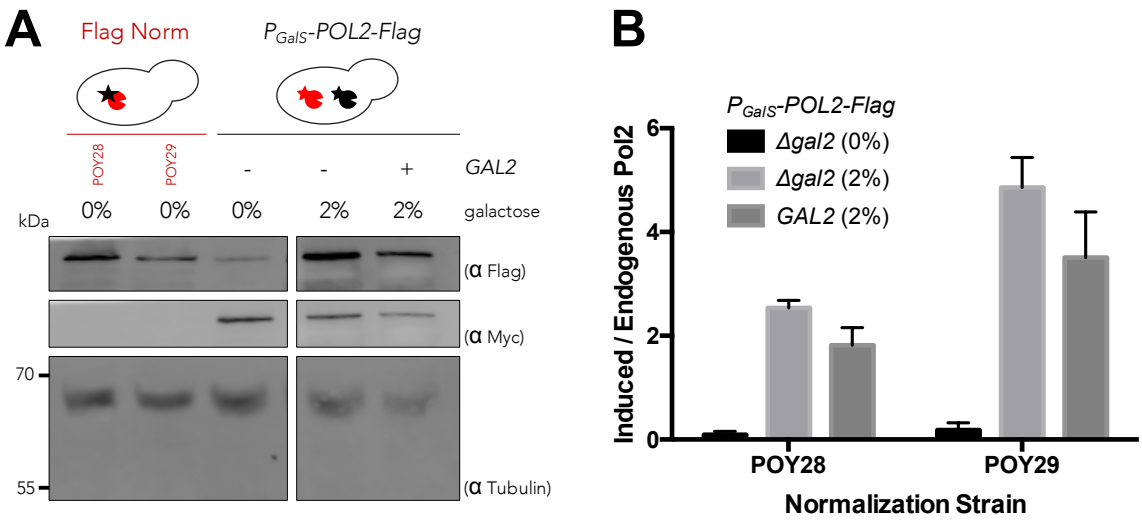



**Figure 45**. **Quantification of ectopic Pol2 induction level**. **(A)** Flag Norm strains (POY28, POY29) were grown in YPD and total protein harvested for a normalization sample. Cultures of *$P_{GalS}$-POL2-Flag;POL2-Myc* (POY25) and *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2* (POY32) were grown in raffinose and either induced with 2% galactose or left uninduced for 2 hours, at which point total protein was harvested and run on a Western blot stained for Flag, Myc, and Tubulin. **(B)** Pol2-Flag signal was quantified with LiCOR software and normalized to the tubulin signal from the same sample. The Induced / Endogenous Pol2 measure was calculated by dividing the tubulin-normalized $P_{GalS}$-Pol2-Flag signal by the tubulin-normalized Pol2-Flag signal from normalization strain POY28 or POY29. Error bars represent a technical repeat of this Western blot.

The approach relies upon an assumption that the ratio of endogenous Pol2 to tubulin is the same in all samples, and that Pol2 stability is not influenced by the

tag. Variability between two Pol2-Flag normalization samples shows that this assumption is not entirely accurate, and that some additional factors (which may include protein stability, or the proportion of the cell population at various stages in the cell cycle) are at play that may vary the cellular ratio of Pol2-Flag to tubulin. However, this approach gives a rough estimate for the galactose induction of the inducible Pol2.

### *Cell cycle impact of Pol2 overexpression*

It was unknown whether overexpression of Pol2 would result in a cell cycle defect. To begin to characterize this, the doubling time of strains with and without the regulatable Pol2 in non-induced (raffinose) or strongly induced (2% galactose) conditions were measured.

The strains all grew similarly in raffinose; the presence of the galactose-inducible Pol2 or the absence of the galactose transporter did not significantly impact cell doubling time under non-inducing growth conditions [Figure 46A]. When grown under galactose induction, however, strains with a galactose-inducible Pol2 showed a statistically significant increase in doubling time compared to the same strain's growth in raffinose [Figure 46B]. This effect was seen with both wildtype Gal2 (*$P_{GalS}$-POL2-Flag;POL2-Myc*) and in the absence of Gal2 (*$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2*), and was not observed in the strains lacking the galactose-inducible Pol2.

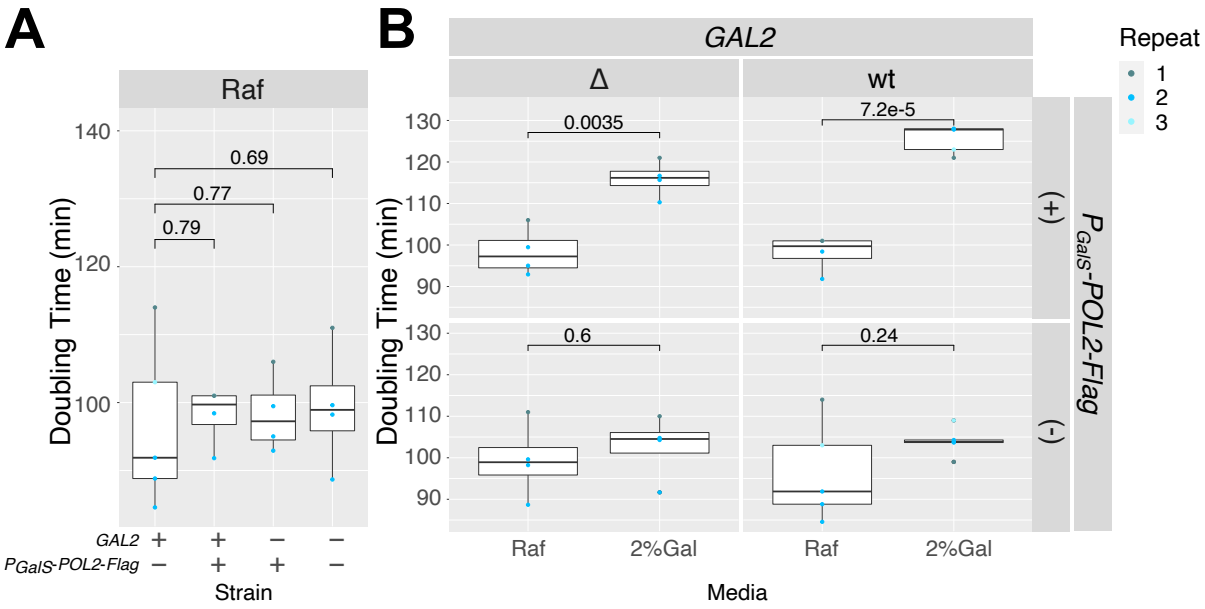



**Figure 46**. **Cell cycle impact of galactose induction of Pol2**. The doubling times of *S. cerevisiae* strains with or without the galactose-inducible Pol2 and galactose-transporter knockout were analyzed in noninducing conditions (raffinose) or under strong induction (2% galactose). **(A)** Comparison of each of the strains (Strain table: POY20, POY25, POY32, POY34) to each other under raffinose growth. **(B)** Comparison of each strain's growth in raffinose versus its growth in 2% galactose. Cell density was measured by spectrophotometry ($OD_{600}$) of cultures growing asynchronously in log phase, and doubling time calculated by fitting an exponential regression to the density over time. Each point represents an independent growth curve, and each 'repeat' is a day in which growth curves were measured. Independent t testing was done between the two media conditions for each strain, and *p* values annotated on the plot.

In addition to observing an increase in the doubling time of strains under Pol2 induction, comparison of the population-wide cell cycle progression of *$P_{GalS}$-POL2-Flag;POL2-Myc* in non-inducing or inducing conditions showed evidence of cell cycle delay [Figure 47]. At 15 minutes after release from pronase, both galactose-induced and non-induced cultures showed a small percentage of the population in G2 (1.56% and 1.58%, respectively). The non-induced population released from arrest and progressed to G2 more rapidly with 17.5% reaching G2 at 30 minutes (versus 7.02% in galactose), 44.5% at 45 minutes (versus 25.3% in galactose), and 61.6% at 90 minutes (versus 44.2% in galactose) [Figure 47B,C].

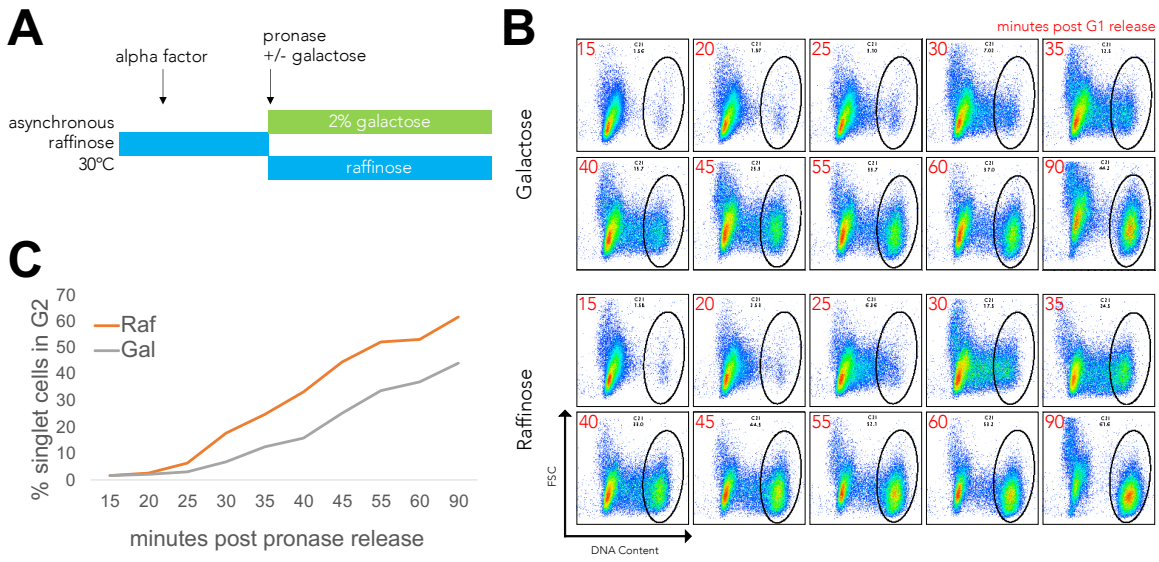



**Figure 47**. **Cell cycle progression defect under Pol2 induction conditions**. **(A)** Schematic for experiment set-up. *$P_{GalS}$-POL2-Flag;POL2-Myc* was grown asynchronously in raffinose at 30°C and arrested with alpha factor. Upon successful arrest, cells were released with 200 µg/mL pronase and the culture split, with one aliquot induced with 2% galactose and one kept in raffinose. **(B)** The cell cycle progression of the cultures upon release from alpha factor arrest was monitored by flow cytometry, with DNA content measured by SYTOX staining plotted against forward light scatter (FSC) as a measure of cell size. The events were gated to remove debris and doublets. The oval shows the gating for the G2 population. **(C)** Progression of singlet cells to G2 after alpha factor release. The percentage of singlet cells within the G2 gate was quantified in FloJo using the gate shown.

### *Titrating Pol2 expression*

Knockout of the galactose transporter, Gal2, has been reported to allow for more sensitive gene expression induction in response to galactose [20]. The ability to tune induction level was anticipated to be valuable both because polymerase switching *in vitro* is sensitive to polymerase concentrations [195], and because the overexpression of Pol2 was associated with a growth defect [Figure 46, Figure 47].

To assess the tunability of galactose induction in the *gal2* knockout strains, the abundance of Pol2-Flag was analyzed in *$P_{GalS}$-POL2-Flag;POL2-Myc* and *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2* induced with 0%, 0.05%, or 2% galactose for 2 hours [Figure 48].

The Gal2 knockout dampened induction in 2% galactose, with Pol2-Flag signal 2-fold higher than the normalization Pol2 in *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2* versus 5.5-fold higher in *$P_{GalS}$-POL2-Flag;POL2-Myc* [Figure 48A]. At 0.05% galactose, Pol2-Flag induction was still high in *$P_{GalS}$-POL2-Flag;POL2-Myc* (2-fold over normalization), but it was barely detectable in *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2* [Figure 48A]. *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2* also showed a change to doubling time relative to *$P_{GalS}$-POL2-Flag;POL2-Myc* in 0.05% galactose. Whereas both *$P_{GalS}$-Pol2-Flag;Pol2-Myc* and *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2* showed significant doubling time delays in 2% galactose relative to raffinose [Figure 46B], in 0.05% galactose *$P_{GalS}$-POL2-Flag;POL2-Myc* showed a significantly increased doubling time (from 118 minutes *POL2-Myc* to 142 minutes *$P_{GalS}$-POL2-Flag;POL2-Myc*) whereas $P_{GalS}$-POl2-Flag;POL2-Myc;Δgal2 showed a less substantial growth defect (from 121 minutes *POL2-Myc;Δgal2* to 127 minutes *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2*) [Figure 48B].

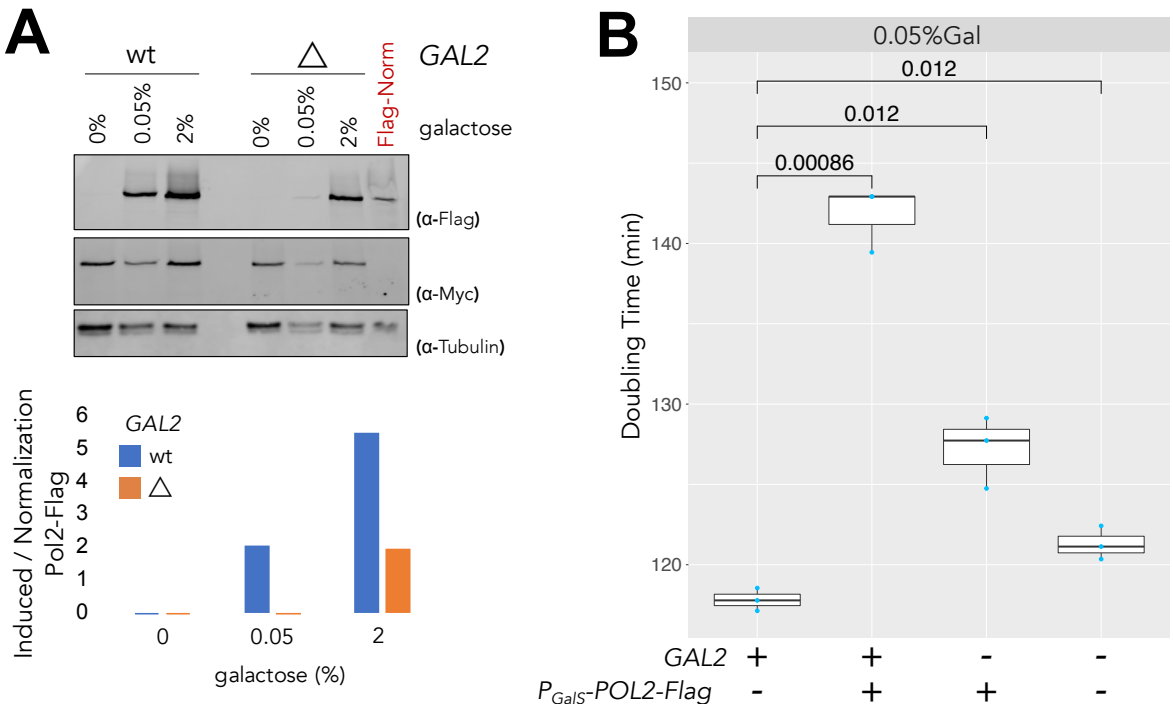



**Figure 48. Titrating Pol2 overexpression in the Gal2 knockout. (A)** *$P_{GalS}$-POL2-Flag;POL2-Myc* (POY25) and *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2* (POY32) were grown asynchronously in raffinose,

0.05% galactose, or 2% galactose at 30°C for 6 hours. Total protein was extracted, and samples were run on a Western blot and stained for Flag, Myc, and Tubulin. Flag signals were normalized to the tubulin signal of the same sample for loading control. Each tubulin-normalized experimental Flag signal was normalized to the signal from the Flag-Norm sample (POY29) to quantify the level of induction relative to the expression level of endogenous Pol2. **(B)** The doubling time of strains with and without *GAL2* and the inducible *$P_{GalS}$-POL2-Flag* was analyzed under conditions of 0.05% galactose induction (in a 2% raffinose YPA media). Cell density was measured by spectrophotometry ($A_{600}$) of cultures growing asynchronously in log phase, and doubling time calculated by fitting an exponential regression to the density over time. Each point represents an independent growth curve. Independent t testing was done between the different strains in this medium, and p values plotted.

A subsequent experiment was carried out to measure the galactose induction response across a range of concentrations. *$P_{GalS}$-POL2-Flag;POL2-Myc* and *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2* were grown asynchronously in raffinose and induced with 0, 0.01, 0.025, 0.05, 0.1, or 0.5% galactose for 2 hours. *$P_{GalS}$-POL2-Flag;POL2-Myc* showed a high level of Pol2 induction in response to low concentrations of galactose, already expressing Pol2 at a higher level than that of endogenous Pol2 in 0.05% galactose [Figure 49]. In contrast, the expression of Pol2-Flag increased more slowly in *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2*, and showed stable expression at about 50% of endogenous Pol2 levels at galactose concentrations of 0.05 and 0.1% before rising to 2.7-fold over endogenous levels at 0.5% galactose [Figure 49].

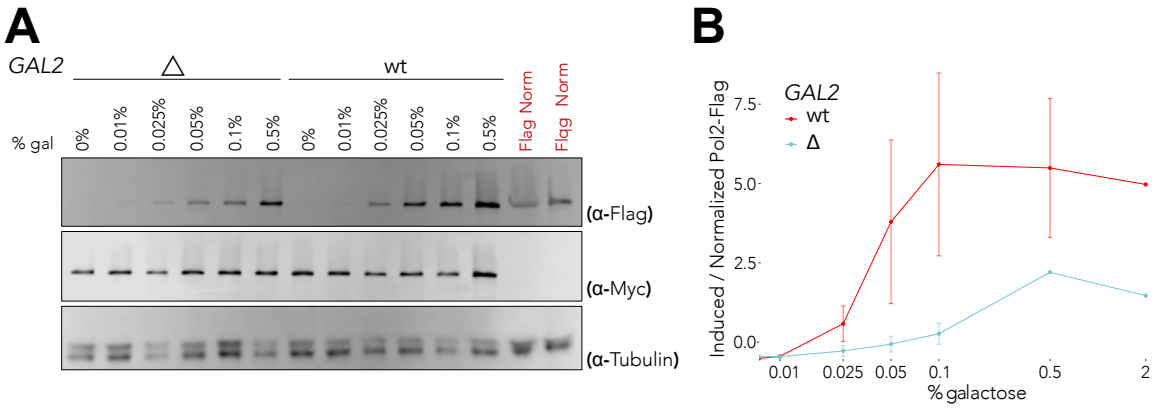



**Figure 49**. **Titrating Pol2 overexpression in the *gal2* knockout**. **(A)** *$P_{GalS}$-POL2-Flag;POL2-Myc* (POY25) and *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2* (POY32) were grown asynchronously in raffinose and induced with galactose concentrations ranging from 0 to 0.5% for 2 hours. Total protein was extracted and Western blotted with antibodies against Flag, Myc, and tubulin. **(B)** Tubulin-normalized induced Pol2-Flag signal over tubulin-normalized normalization Pol2-Flag signal

quantification from this Western, a technical repeat of these samples, and a biological repeat. The error bars plot standard error of the measurements.

### *Population heterogeneity in galactose induction of Pol2*

Using Western blotting to determine galactose induction yields valuable insight into the ensemble response of the population but may hide heterogeneity that exists within the population. This raised the question of whether the dampened galactose induction in the galactose transporter knockout strain was the result of homogenously-decreased expression across the entire population, or rather a substantial expression decrease in some of the population, but little to no change in other portions of the population. To address this question, a single-cell method for assaying galactose-induced Pol2 expression was required.

To achieve this, a plasmid was designed to introduce Pol2 tagged with a red fluorescent protein Ruby under the GalS promoter at the *HO* locus. This would enable analysis of Pol2 expression in single cells by flow cytometry – both allowing for identification of population heterogeneity in galactose induction response and facilitating work to investigate any potential impact of polymerase overexpression at various stages of the cell cycle.

Work was paused at this point due to the coronavirus.

## Cell cycle switching experiments

### *Arrest in HU following constitutive induction or repression*

With the inducible Pol2 system in place and an effective cell synchronization protocol (although before *gal2* knockout or the double spike-in creation), it was possible to design experiments to test for switching of the induced Pol2 onto chromatin during ongoing replication. Cell arrest with hydroxyurea (HU) is known to allow for the specific firing of subsets of early-firing origins ('unchecked' origins), but the replication fork progression from these origins is limited as the replication checkpoint is quickly activated in response to depleted nucleotide pools [173]. This yields a strong expectation for Pol2 peak locations, and thus poses a good system for testing that the induced Pol2 can be detected on the chromatin and for quantifying the sensitivity of the ChIP-Seq protocol.

Therefore, an experiment was carried out in which a *$P_{GalS}$-POL2-Flag;POL2-Myc* culture with constitutive Pol2 expression (2% galactose) and one with constitutive repression (glucose) were both released from alpha factor arrest into HU [Figure 50A]. After 90 minutes in HU, cells from each sample were harvested, spiked with *S. bayanus* Flag IP spike-in cells (Strain Table: CAY300), then cross-linked and carried though the ChIP-Seq protocol for Flag and Myc (see Methods). Additionally, samples were taken for flow cytometry at the start of the experiment, at the point of alpha factor arrest before release with pronase, and at 90 minutes in HU [Figure 50B]. Flow cytometry showed a successful arrest of both samples, but at 90 minutes after release into HU the repressed sample showed more evidence of progression from G1 and DNA content increase than the galactose-induced sample. A total protein sample for Western blotting was taken after 90 minutes in HU, in parallel to the ChIP-Seq and flow samples.

The sequenced reads from each library were normalized for total library sequencing depth and then calibrated based upon the spike-in for that library. As the cells were arrested in HU there was a strong expectation of signal spikes flanking unchecked origins. Previously-published BrdU-IP data from HU-arrested cells were used confirm the genomic locations and extent of DNA replication expected [44].

The expectation of Pol2 peaks flanking unchecked origins was met in the endogenous Pol2-Myc signal (Myc IP) from the inducible-Pol2 repressed sample: the Pol2 Myc reads formed peaks surrounding the genomic regions known to be replicated before HU arrest [Figure 50E]. As expected, the repressed sample did not show peaks for the inducible Pol2-Flag ChIP – this is in keeping with Western blot data showing no evidence of Pol2-Flag expression in the repressed sample [Figure 50C].

In contrast to the repressed sample, the read traces for the induced sample did not match expectation. There was no evidence of peaks flanking unchecked replication origins for either the endogenous or inducible Pol2 in the induced sample [Figure 50E].

To assess these findings across replication origins systematically, the calibrated read counts were pulled for 75 kb windows flanking each checked or unchecked replication origin. The average and 95% confidence intervals for these read counts were plotted for each sample [Figure 50D]. The endogenous Pol2-Myc IP from the repressed sample showed the highest read counts across the board and read count peaks were located 25 kb on either side of origins.  The checked

origins in the repressed sample showed no read enrichment either directly at or flanking the origins [Figure 50D]. This is in good agreement with the expectation.

The induced sample interestingly showed a minor peak of both endogenous and inducible Pol2 signal centered on checked origins – this same minor peak was not evident at unchecked origins [Figure 50D].

Based upon these data, it appeared that the Pol2-Flag repressed sample had, as hoped, released from G1 arrest and been arrested at early S phase. DNA content had increased from G1 as indicated by flow cytometry, and the endogenous polymerase epsilon was found on chromatin at the expected arrest sites. There was no evidence of the regulatable Pol2-Flag on the chromatin in the repressed sample. The induced sample, however, showed no evidence of release from G1 based on bulk DNA content assayed by flow cytometry, nor was the endogenous or ectopic induced polymerase epsilon found on the chromatin at sites flanking unchecked replication origins. Together, this is consistent with the constitutive overexpression of the polymerase epsilon Pol2 subunit impeding firing of early origins under conditions of replication stress induced by hydroxyurea. This is far from conclusive, as this experiment does not explicitly have a positive control for the Flag IP. However, the fact that the endogenous Pol2 reads for the induced sample show the same behavior as the inducible Pol2 reads for the induced sample – rather than echoing the endogenous Pol2 reads of the repressed sample – suggests that the problem lies in the cell cycle status or viability of the induced sample, rather than in the Flag IP itself.

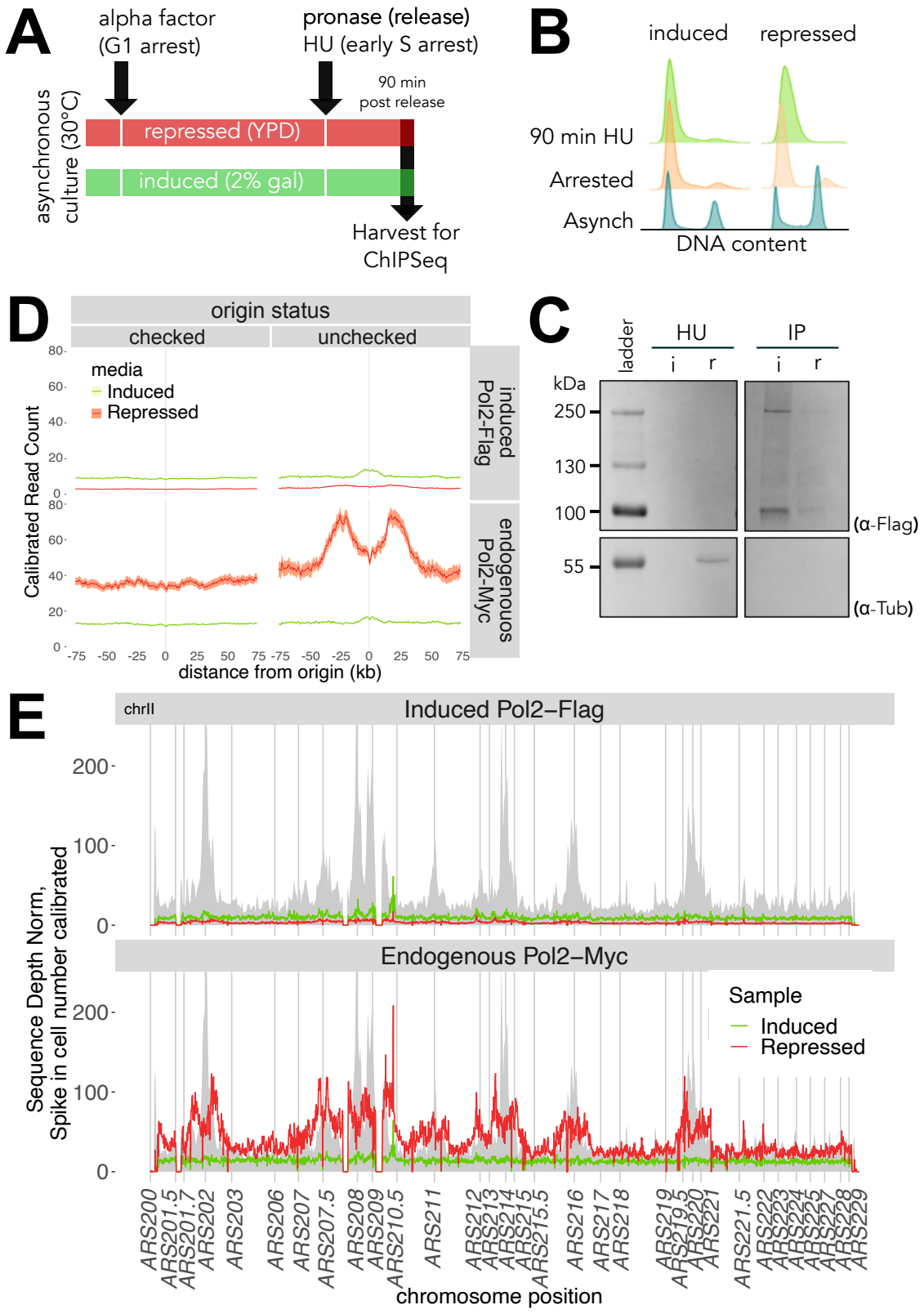



**Figure 50**. **Arrest in HU following constitutive induction or repression**. **(A)** Schematic for experiment design. *$P_{GalS}$-POL2-Flag;POL2-Myc* (POY25) grown asynchronously in repressive (YPD) or inducing (2% galactose) conditions was submitted to an alpha factor arrest. Upon successful G1 arrest confirmed by microscopy, cells were released from G1 and arrested in early S phase with the simultaneous addition of 200 µg/mL pronase and 200 mM HU. Cells were harvested for ChIP-Seq 90 minutes after the addition of pronase and HU. **(B)** Cell cycle was

monitored with flow cytometry at the onset of the experiment (Asynch), at the point of alpha factor arrest (Arrested) and 90 minutes after the addition of pronase and HU. **(C)** Samples were taken from both induced (i, 2% galactose) and repressed (r, YPD) cultures for total protein extraction after 90 minutes in HU in parallel to taking samples for ChIP-Seq. Both the 90 minute HU whole cell extract samples (HU) and samples that had been processed through the Flag IP (IP) were run on a Western blot and probed for Flag and tubulin. The induced IP sample showed evidence of Flag-tagged polymerase, while the repressed IP sample did not. **(D)** Calibrated read counts covering 75 kb on either side of confirmed origins of replication were captured for each dataset, and origins were classed into 'checked' (those that are not active under HU) versus 'unchecked' (those that are active under HU) following [44]. The plot shows average read count and 95% confidence intervals as a ribbon. **(E)** Visualization of calibrated read counts across chromosome II for Flag (Induced Pol2) and Myc (Endogenous Pol) IPs from both repressed and induced samples. Origin locations are plotted along the X axis. BrdU-IP reads from cells arrested in HU published in (Müller et al., 2019) are plotted in grey for comparison.

### *Arrest in HU with non-induction or induction*

As the data suggested that constitutive overexpression of Pol2 had impeded either successful replication initiation upon release from HU or fork stabilization under HU arrest, an experiment was conducted to delay Pol2 induction until the point of alpha factor release into HU [Figure 51A]. A culture of cells was grown asynchronously in raffinose, arrested with alpha factor, and split upon pronase release into HU such that one aliquot remained uninduced in raffinose while the other was induced with 2% galactose. Samples were taken for both flow cytometry and Western blotting at the start of the experiment, at alpha factor arrest just prior to pronase release, and after 90 minutes in HU [Figure 51B,C]. Based on the flow cytometry profiles, both the uninduced and induced samples had undergone a subtle increase in DNA content at 90 minutes following alpha factor release into HU [Figure 51B]. Western blotting confirmed the absence of induced Pol2-Flag in the asynchronous and alpha factor arrested samples, and its presence in the galactose-induced sample after 90 minutes in HU, as expected [Figure 51D]. Surprisingly, the raffinose sample at 90 minutes in HU showed some Pol2-Flag signal [Figure 51D]. Raffinose is known to be non-inducing of the GalS promoter (in keeping with the absence of Pol2-Flag bands in previous raffinose samples, including the asynchronous and alpha factor arrested samples), but with

some leakiness. The appearance of the Pol2-Flag band specifically under raffinose arrest in HU suggests that this leakiness may be accentuated or normal levels of Pol2 turnover inhibited by the HU arrest or the HU-associated replication checkpoint.

The ChIP-Seq reads for the endogenous Pol2 (Myc) localized exactly as expected, with both the induced and non-induced samples showing peaks flanking early firing unchecked origins [Figure 51E]. The peak height was slightly higher for the induced sample than the non-induced, which runs contrary to the expectation that the endogenous Pol2-Myc might be diluted from the chromatin by switching to the induced Pol2-Flag [Figure 51E]. More surprisingly, in the Flag IP for the induced Pol2, the non-induced sample behaved like the induced sample, with peaks at known loci of HU arrest. This appearance of induced Pol2 ChIP-Seq signal in the non-induced sample is in keeping with the Western blot; however, the peak heights of the non-induced sample were higher than those of the induced sample, which was contrary to the stronger Western signal in the induced sample than the non-induced [Figure 51 C,E].

A straightforward explanation for this would be that the samples had been switched at some point in the library-making process. This possibility was addressed with an approach that exploits the fact that actively transcribed regions are known ChIP-Seq artifacts [209]. Briefly, in the galactose-induced sample, the ectopic Pol2-Flag under the control of the GalS promoter is expected to be expressed highly. In addition to this, the galactose genes (including *GAL1*, *GAL2*, *GAL7*, and *GAL10*) are coordinately activated in response to galactose. The transcriptional activation of these genes in the galactose-induced sample is

expected to drive more ChIP-Seq pull-down, leading to a higher read count over these loci. Analysis of reads over the Gal genes and reads mapped to the *POL2* gene showed higher read counts in the induced sample, consistent with the samples not being switched [Figure 51D].

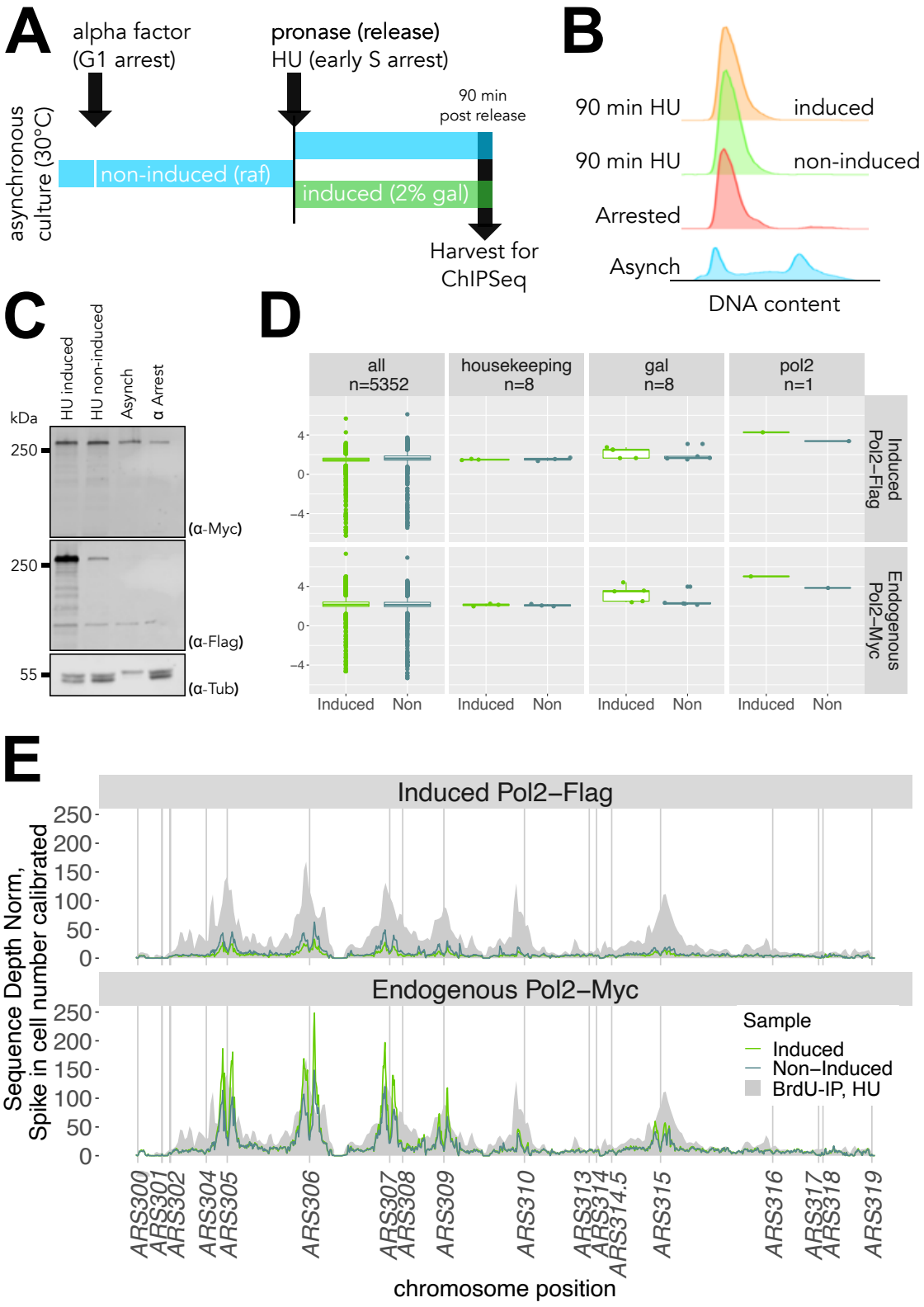



**Figure 51**. **Arrest in HU in inducing or non-inducing conditions**. **(A)** Schematic for experiment design. A *$P_{GalS}$-POL2-Flag;POL2-Myc* (POY25) culture grown asynchronously in non-inducing medium (raffinose) was submitted to alpha factor arrest. Upon successful arrest, the culture was split into two aliquots; both were released from G1 and arrested in early S phase with the simultaneous addition of 200 µg/mL pronase and 200 mM HU. One aliquot was induced with

galactose to a final concentration of 2%. Cells were harvested for ChIP-Seq 90 minutes after the addition of pronase and HU, +/- galactose. **(B)** Cell cycle progression was monitored by flow cytometry for DNA content (SYTOX) at the start of the experiment (Asynch), at the point of alpha factor arrest (Arrested) and 90 minutes after the addition of pronase and HU for both the induced and non-induced samples. **(C)** Total protein samples were taken for Western blotting at the onset of the experiment (Asynch), at the point of alpha factor arrest (α Arrest), and after 90 minutes in HU. **(D)** Boxplots of the mean calibrated read count for IP samples was calculated across regions annotated for (1) all genes (2) a set of housekeeping genes (3) genes of the galactose metabolic pathway and (4) Pol2. **(E)** Visualization of calibrated read counts across chromosome III for Flag (Induced Pol2) and Myc (Endogenous Pol2) IPs from both induced and non-induced samples. Origin locations are plotted along the X axis. BrdU-IP reads from cells arrested in HU published in (Müller et al., 2019) are plotted in grey for comparison.

### *S phase progression with non-induction or induction*

Given the unexpected accumulation of Pol2-Flag signal in the non-induced sample under HU arrest, the next experiment avoided HU-induced replication stress and assessed the chromatin localization of Pol2 in induced or non-induced cells released from alpha factor arrest to progress freely through S phase [Figure 52A]. Samples were taken for flow cytometry after pronase release from G1 arrest to monitor progression through S phase in the galactose-induced and non-induced samples [Figure 52B]. Chromatin was harvested for ChIP-Seq at 30 and 45 minutes following pronase release [Figure 52A].

Western blot analysis of protein extract from the samples at 30 minutes and 45 minutes again showed an absence of inducible Pol2-Flag in the original asynchronous culture or the culture at the point of alpha factor arrest [Figure 52C]. Calibrated read counts for the Flag and Myc IP across gene sets showed higher pull down of genes associated with galactose metabolism and the Pol2 gene in the induced samples at 30 and 45 minutes than in the non-induced samples, providing further confirmation that the induction had had the expected impact on gene expression [Figure 52D]. The traces for induced Pol2-Flag and endogenous Pol2-Myc IP reads did not show peaks as expected for well-synchronized progression through S phase [Figure 52E]. This fits with the flow cytometry data

showing relatively uncontrolled progression from the G1 arrest to G2 [Figure 52B]. Unfortunately, the low level of synchrony and the resultant lack of structure to the IP traces makes this dataset of limited use for identifying polymerase epsilon switching *in vivo,* as it limits the ability to discriminate real IP signal from background noise.

To use these data to get a broader indication of whether the experimental set-up is conducive to assaying for polymerase switching (given better control of S phase synchrony), general trends of IP reads across replication timing bins were assayed. As early-replicating regions were expected to have completed replication before the induction of the Pol2-Flag, it was expected that the Flag-IP of induced samples would show higher IP read coverage in later-replicating regions of the genome. Additionally, as the induced Pol2-Flag had more time to express and accumulate in the 45 minute induced sample than the 30 minute induced sample, it was expected that calibrated Flag IP reads would be higher across the board for the 45 minute sample.

These expectations were tested by making use of replication timing data for the parental strain VGY129 generated by A. Wallis (personal communication). The genome was split into replication timing deciles (from those genomic loci replicated in the first tenth of S phase to the last tenth) and the count of calibrated Flag-IP reads for bins classed into each of these deciles was plotted [Figure 53]. Both the expectation of increased calibrated read counts in bins of later replication timings and the expectation of overall higher calibrated read counts for the 45 minute sample were met, supporting this experimental set-up as a system with which to assay polymerase switching, if improved synchrony could be achieved.

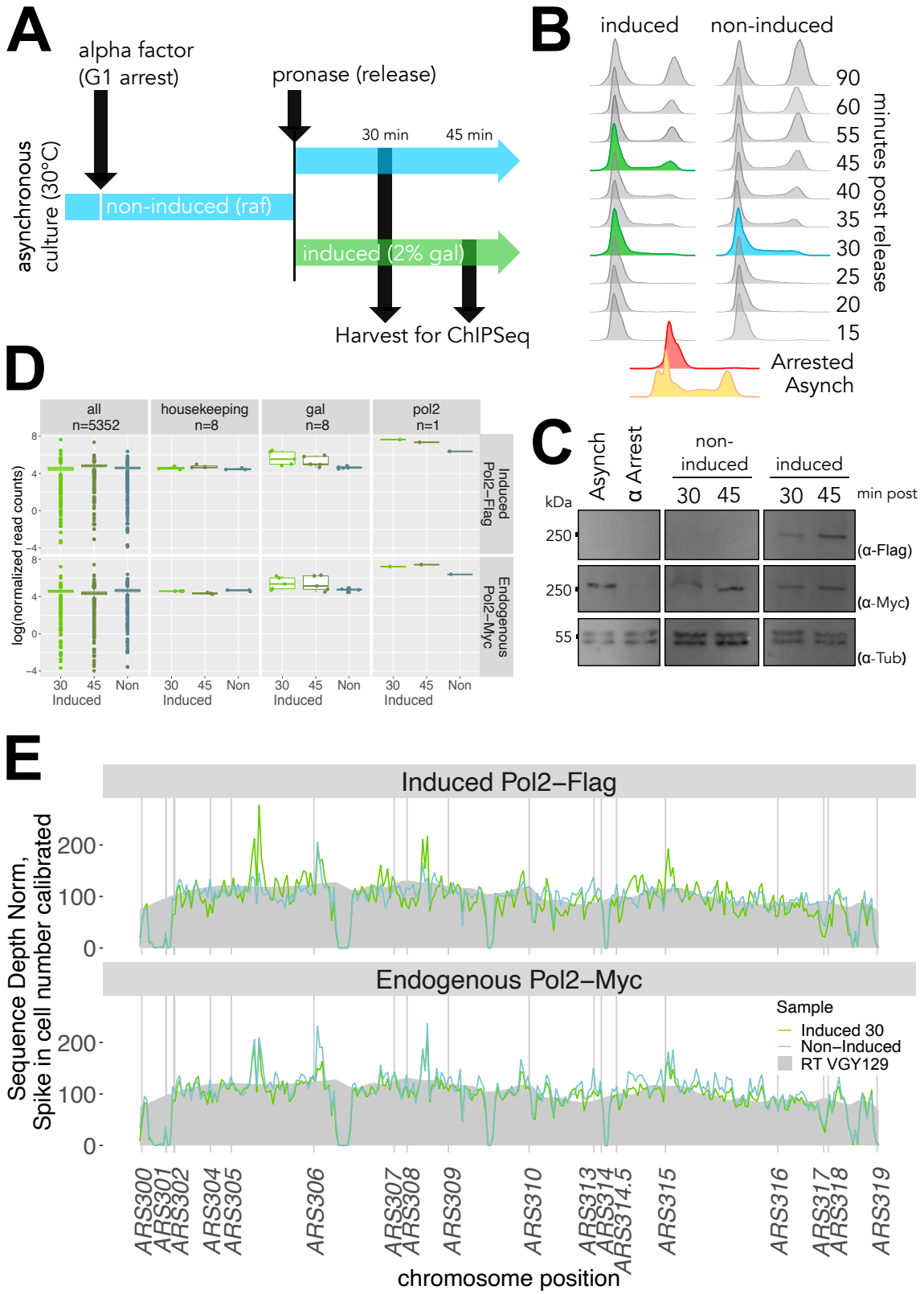



**Figure 52. S phase progression in inducing or non-inducing conditions. (A)** Schematic for experimental design. A *$P_{GalS}$-POL2-Flag;POL2-Myc* (POY25) culture was grown asynchronously in non-inducing media (raffinose) and submitted to alpha factor arrest. Upon successful arrest, the culture was split into two aliquots; both were released from G1 with the addition of 200 µg/mL pronase. One aliquot received induction with galactose to a final concentration of 2%. Cells were harvested for ChIP-Seq at 30 and 45 minutes following the addition of pronase +/- galactose. **(B)**

Cell cycle progression was monitored by flow cytometry at the start of the experiment (Asynch), at the point of alpha factor arrest (Arrested), and at intervals following pronase release for both the induced and non-induced samples. **(C)** Total protein samples were taken for Western blotting at the start of the experiment (Asynch), at the point of alpha factor arrest (Arrest), and at 30 and 45 minutes following pronase release from G1. **(D)** Boxplots of the mean calibrated read count for IP samples were calculated across regions annotated for (1) all genes (2) a set of housekeeping genes (3) genes of the galactose metabolic pathway and (4) POL2. **(E)** Visualization of calibrated read counts across chromosome III for Flag (Induced Pol2) and Myc (Endogenous Pol2) IPs from both induced and non-induced samples at the 30 minute time point. Origin locations are plotted along the X axis. Replication time calculated from SortSeq data from A. Wallis in parental strain VGY129 is plotted in grey for comparison.

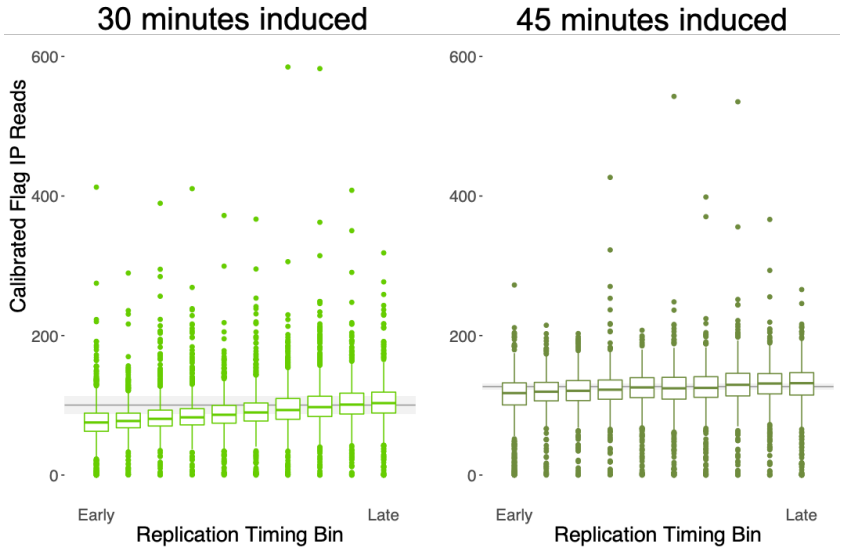



**Figure 53. Flag-IP reads in galactose-induced 30 minute and 45 minute samples.** Replication timing (RT) data from A. Wallis (personal communication) were used to assign RT deciles (from early to late replicating) to every bin in the genome. The calibrated Flag-IP read counts for bins in each RT decile was plotted in boxplots for the 30 minute and 45 minute induced samples. In grey ribbon is the mean and 95% confidence interval for calibrated read count across the genome as a reference for the change across the deciles.

### *Slowed S phase progression (23°C) with non-induction or induction*

To improve synchrony, the arrest-release time-course was repeated at 23°C. As before, cells were grown asynchronously in raffinose, arrested with alpha factor, and released with pronase into two aliquots. One was induced with 2% galactose and one left uninduced [Figure 54A]. Samples were taken for flow cytometry over the release period [Figure 54B] and for ChIP-Seq and Western blotting at 40 minutes and 60 minutes following pronase release. The synchrony upon pronase release was improved over that seen for growth at 30°C, however again the non-

induced culture proceeded marginally more quickly through S phase than the induced culture [Figure 54B].

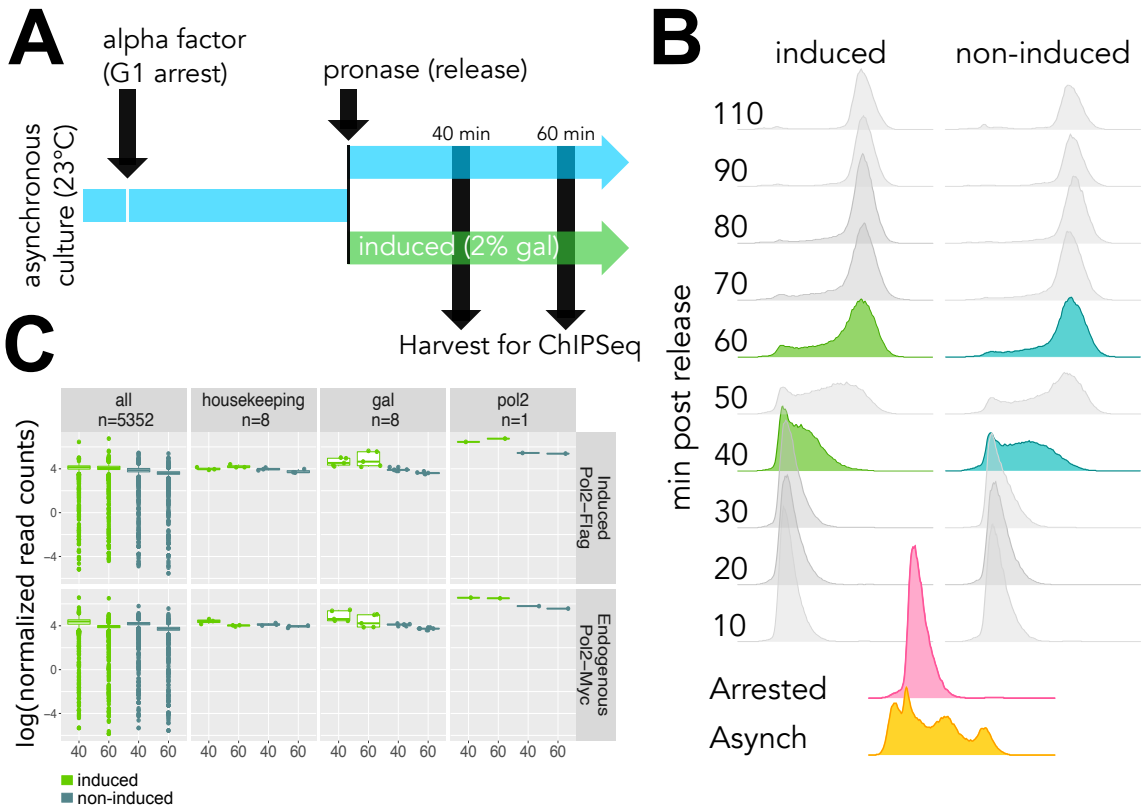



**Figure 54. Slowed S phase progression at 23°C in inducing or non-inducing conditions. (A)** Schematic for experimental design. A *$P_{GalS}$-POL2-Flag;POL2-Myc* (POY25) culture was grown asynchronously at 23°C in non-inducing media (raffinose) and submitted to alpha factor arrest. Upon successful arrest, the culture was split into two aliquots; both were released from G1 with the addition of 200 µg/mL pronase. One aliquot received induction with galactose to a final concentration of 2%. Cells were harvested for ChIP-Seq at 40 and 60 minutes following the addition of pronase +/- galactose. **(B)** Cell cycle progression was monitored by flow cytometry at the start of the experiment (Asynch), at the point of alpha factor arrest (Arrested), and at intervals following pronase release for both the induced and non-induced samples. **(C)** The mean calibrated read count for IP samples was calculated across regions annotated for (1) all genes (2) a set of housekeeping genes (3) genes of the galactose metabolic pathway and (4) POL2.

Analysis of the ChIP-Seq reads showed both induced and endogenous Pol2 peaks at locations anticipated based upon beacon calculus simulations of polymerase progression over S phase [Figure 55A]. In the 40 minute timepoint, endogenous Pol2-Myc reads predominated in both the induced and non-induced sample. By 60 minutes following induction, in the non-induced sample the endogenous Pol2-Myc reads still predominated, however in the induced sample at

60 minutes the induced Pol2-Flag IP was at higher abundance than the endogenous Pol2-Myc IP reads, supporting the notion that at 60 minutes post galactose-addition, the inducible Pol2-Flag had accumulated in the induced sample and was localizing on the chromatin [Figure 55A].

To assess whether the presence of the induced Pol2-Flag on the chromatin was the result of polymerase switching into ongoing replisomes or polymerase introduction into newly initiating replisomes, the relationship between the abundance of induced Pol2-Flag reads and the characteristic replication fork age of a given genomic locus was assessed [Figure 55B]. As seen through analysis of IP traces over chromosomes, the induced 60 minute sample was the only sample which showed any evidence of higher abundance of induced Pol2-Flag reads than endogenous Pol2-Myc reads. Interestingly, this enrichment of induced Pol2 reads over endogenous Pol2 reads was most pronounced in genomic bins with replication fork ages of less than 25 minutes. At genomic loci on average replicated by forks older than 25 minutes old, the difference in induced Pol2-Flag versus endogenous Pol2-Myc reads became less distinct.

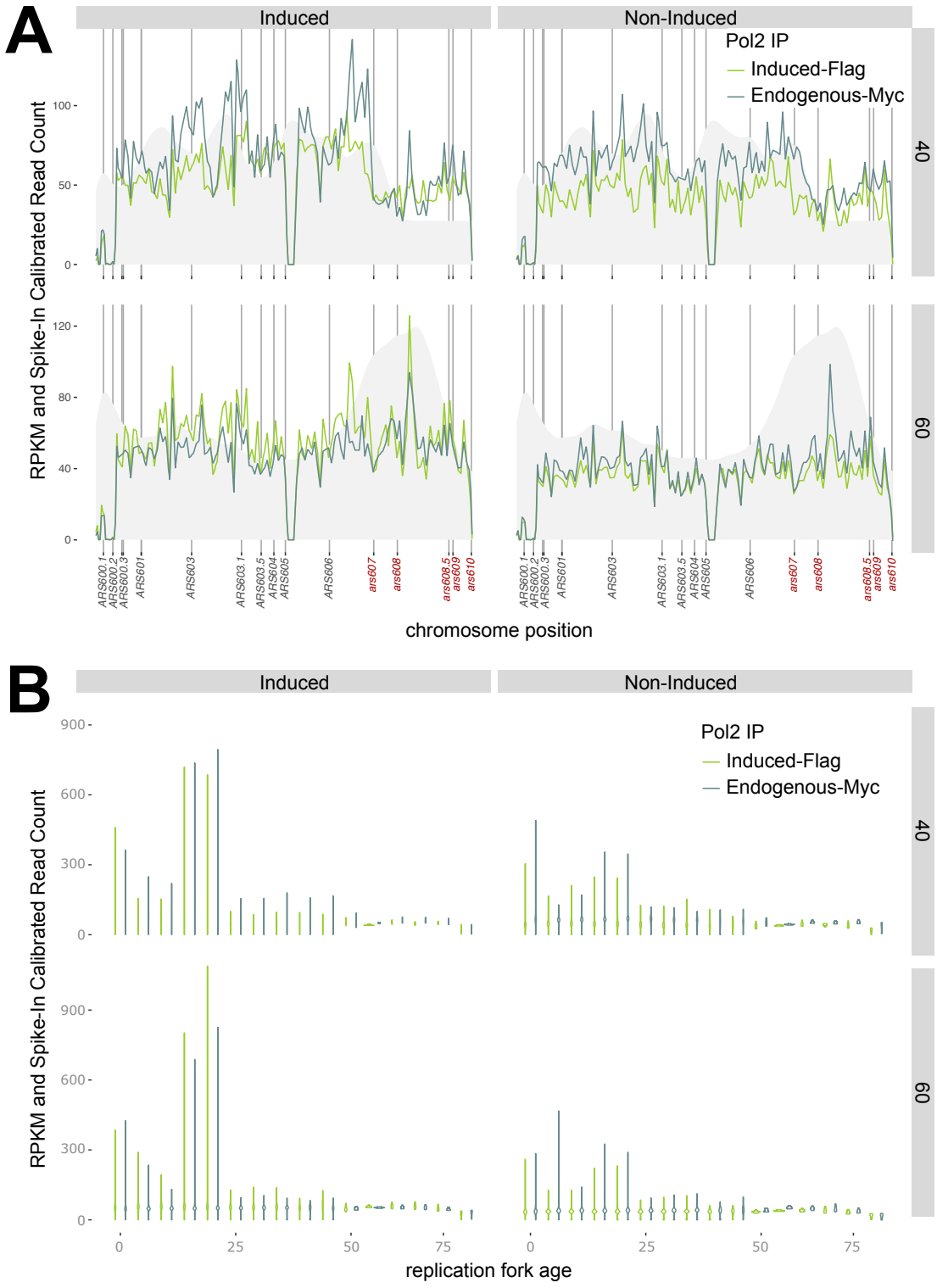



**Figure 55. Slowed S phase progression at 23°C in inducing or non-inducing conditions. (A)** Visualization of calibrated read counts across chromosome VI for Flag (Induced Pol2) and Myc (Endogenous Pol2) IPs from both induced and non-induced samples at the 40 minute (top) and 60 minute (bottom) time points. Origin locations are plotted along the X axis, with inactivated ori in the experimental strain shown in red. The expected polymerase positions for the given S phase position as predicted by beacon calculus simulations are plotted in grey. **(B)** Induced and endogenous polymerase IP versus replication fork age. The median fork age for each genomic bin

was calculated from beacon calculus simulations. Calibrated read counts across the genome were binned by the fork age of the genomic bin and plotted as a violin plot.

## Discussion

Switching of the leading strand polymerase during DNA replication may offer flexibility to ensure the safeguarding of genome stability in response to replication stress or template damages. Enzyme switching and recycling are widely understood to be integral to lagging strand replication, as the synthesis of each Okazaki fragment depends upon novel engagement of a polymerase delta complex with a new primed fragment. In contrast, the model of a singly-primed leading strand opens the possibility for continuous synthesis by one specific polymerase epsilon complex directly tracking with the progression of helicase unwinding. This possibility seemed particularly compelling given early protein structure findings of tight interactions between polymerase epsilon and the CMG [210].

However, this simple model of tight polymerase epsilon conjugation to the CMG and continuous leading strand replication leaves large questions about how the system might respond to replication stalls, template obstacles, and other complexities inherent to *in vivo* replication. In beginning to address those questions, *in vitro* work has shed important light on the dynamic nature of the replisome machinery. Biochemical studies have shown potential for replication restart following blockage by a leading strand obstacle [188], and microscopy studies have expanded on this to show evidence of polymerase epsilon switching during ongoing unperturbed replication [195]. While the *in vitro* work offers invaluable opportunities to gain mechanistic insight through precisely-defined experimental conditions, *in vivo* work is required to further understand the physiological relevance of polymerase switching and dynamic replisome behavior.

Additionally, sequencing-based work offers the opportunity to look at site-specific behaviors, such as programmed fork arrest or transcription conflict. Thus, this work has endeavored to create a yeast system allowing for *in vivo* investigation of polymerase switching during ongoing leading strand replication.

### Creation of a regulatable polymerase system in *S. cerevisiae*

In this work, two methods to create a regulatable polymerase epsilon Pol2 subunit were tested. The first employed a post-translational localization control approach, by fusing an estradiol hormone binding domain tag (E tag) onto the Pol2 C terminus. While this E tag has been previously shown to be effective in driving cytoplasmic sequestration of proteins of interest [199], in this application the E tag was insufficient for localization control of Pol2. This could be due to Pol2 expression levels exceeding the cellular pool of HSP proteins required for cytoplasmic sequestration. Previous work using E tag sequestration applied this method to a transcription factor under the constitutive low-to-moderate expression of *pACT* [199], rather than the native Pol2 promoter which shows cell-cycle dependent expression peaking in S phase. If saturation of the HSP proteins caused the observed nuclear leakiness of the Pol2-E tag protein, it would be possible to improve this approach utilizing a double-induction method, either by putting the Pol2 protein under an inducible promoter that can be turned on simultaneously to the addition of estradiol, or by adding an inducible degron tag and stopping degradation simultaneously to the addition of estradiol.

Another possibility is that the size or conformation of the tagged protein interferes with the E tag sequestration. The Pol2 protein, at 255 kDa, is substantially larger than the 52.5 kDa (including the E tag) transcription factor fusion protein studied

in McIsaac *et al*. Furthermore, Pol2 exists in the polymerase epsilon complex. It is possible that the hormone binding domain tag is buried in the Pol2 folds or within the polymerase epsilon complex, thus precluding sequestration. Fusing the tag to the N terminus or at different internal sites along the Pol2 protein might help to ensure that it is accessible for sequestration.

Unlike the estradiol tag localization control approach, transcriptional control proved to be very effective for creating a regulatable Pol2 system. The Pol2 protein was not detected via Western blotting in repressive glucose growth conditions, nor under non-inducing raffinose growth conditions except for when HU was present. Upon addition of 2% galactose, the Pol2 protein was induced to levels 2 to 5-fold higher than the endogenous Pol2 protein within 2 hours. This system was rapid, with the protein first being detected by Western blotting 20 minutes after galactose induction, and tunable, with high levels of Pol2 induction tamped by using lower galactose levels in a Gal2 knockout background. ChIP-Seq experiments showed evidence for potential leakiness of this system at levels below the detection limit of Western blotting. To further improve on this system and reduce the potential for background leakiness, it would be useful to add an inducible degron tag to the inducible protein. With this addition, in the non-induced transcriptional state, any expression leakiness could be mopped up by protein degradation, and upon transcription induction the degradation could simultaneously be turned off.

### Cell impacts of polymerase epsilon overexpression

Creating the inducible polymerase expression system gave the opportunity to gain insight into the cellular ramifications of Pol2 overexpression. Interestingly,

overexpression of Pol2 was shown to result in a cell cycle defect not previously reported, with a longer cell doubling time as measured in asynchronous cultures and delayed progression through S phase assayed in synchronized cells upon release from alpha factor.

While in-depth investigations into the cellular mechanisms driving this defect were outside of the scope of this work, the finding reemphasizes the multiple roles that polymerase epsilon plays in S phase function. While overexpression of Pol2 could threaten the stoichiometry of the polymerase epsilon complex by titrating away other complex subunits, it is suspected that neither titration of Dbp3 nor Dpb4 would cause this phenotype as both subunits are non-essential. Interaction with the essential subunit Dpb2, however, could result in replication delays if free Pol2 could pull Dpb2 away from its inclusion in a polymerase epsilon complex conjugated to the CMG, as Dpb2 has been shown to be critical for processive synthesis [159].

In addition to potential interference with Dpb2, another intriguing possibility for the defect is that Pol2 overexpression may titrate Mrc1 away from active replisomes. Mrc1 interacts independently with the N and C termini of Pol2, and the interaction between the two proteins is strong enough to be observed through immunoprecipitations without crosslinking and with salt washes [139]. In wildtype cells, Mrc1 was found to exist in a 2:1 stoichiometry to Pol2, and all cellular Pol2 was found to immunoprecipitate with Mrc1 [139]. Interestingly, Mrc1 is required both for maximum replication fork speeds [180] and for Rad53 activation in response to replication stress [172], and thus impacts on Mrc1 localization could account for both

slowed cell cycling in unperturbed cells and failure to initiate DNA replication under HU treatment in Pol2 overexpression cultures.

Further work to investigate the impact of Pol2 overexpression on the localization and replisome incorporation of Dpb2, Mrc1, and other replisome components will be interesting to test the hypothesis that Pol2 overexpression drives genome instability by titrating fork components away from active replication forks. A useful initial test could be a simple rescue study introducing overexpression of Dpb2, Mrc1, and other candidate components to test whether the cell cycle defect induced by Pol2 overexpression can be ameliorated with the commensurate overexpression of Pol2 interactors.

As the mechanism is not fully understood, the observation of cell cycle defects in the Pol2 overexpression system raises the caveat that findings in this system must be understood to be achieved from an *in vivo* system that still does not perfectly represent wildtype *in vivo* replication conditions. *In vitro* work has shown that free polymerase concentration impacts switching behavior independent of any replication checkpoint functions [195] – it is possible that perturbations to Pol2 abundance *in vivo* could result both in this concentration-dependent impact on switching behavior and in pleiotropic impacts on replication through interactions with the CMG and related proteins.

### Assaying for switching behavior

The ChIP-Seq experiments assaying over cell progression through S phase offer the first opportunity to assay in eukaryotes for evidence of polymerase switching *in vivo*.

The first experiment pulling down endogenous and induced Pol2 in cells synchronized and arrested in HU showed the effectiveness of this system for identifying the genomic locations of the leading strand polymerase on the chromatin. The second experiment, in which the Flag-tagged Pol2 was induced concomitant to the release of cells from alpha factor and their arrest in HU, showed the functionality of the newly induced polymerase for incorporation into the replication fork.

The following two experiments, which released *S. cerevisiae* into S phase progression with Pol2 induction, allowed for evaluation of polymerase switching into ongoing forks. A powerful way to test the ability of an induced Pol2 to switch into an ongoing replisome is to assess the presence of the induced Pol2 at genomic regions which have, on average, old replication forks versus regions which are replicated, on average, by newer forks. If the potential for Pol2 switching into an ongoing replisome is commensurate to its potential to be incorporated into a newly initiated fork, the abundance of induced Pol2-Flag reads should be equivalent at sites with older forks versus those with newer forks for a given replication timing window. The ultra-long replicon strain used in this work offers a particularly powerful platform for assessing this distribution of induced Pol2 at old and young forks as the origin deletions on the right arm of chromosome VI create a replicon of about 100 kilobases, necessitating continued progression of a single replisome for an average of 75 minutes – substantially longer than the persistence of an average replisome on the chromatin.

Across the genome these experiments hinted at a higher abundance of induced Flag-tagged Pol2 in regions synthesized by young replication forks than those

synthesized by older forks – such as those found on the ultra-long replicon on the right arm of chromosome VI. This is consistent with an absence of common and rapid polymerase switching during ongoing replication. These experiments merit replication to validate and better understand the extent of the potential bias of induced polymerase towards new forks. Such experimentation would also benefit from a system that allows for more instances of long replicons and resulting old replication forks genome-wide (such as the *cdc7-ts* system).

A lack of evidence for widespread polymerase switching in these experiments does not, however, preclude the possibility that switching occurs *in vivo*. Further experiments using perturbations hypothesized to induce switching behavior will be interesting to push forward understanding of the potential role for polymerase switching. As in vitro work has shown that increased Pol2 abundance promotes switching behavior [195], higher over-expression of the induced Pol2 may be interesting, as would the introduction of replication stress in the form of UV-induced DNA damage.

## Future directions

### *Alternative read-out*

Using a differential tag on the induced polymerase allowed for identification of polymerase localization via ChIP-Seq. While very powerful, this approach has two potential weaknesses.

First, this approach yields information on the likelihood of the tagged polymerase to crosslink with and immunoprecipitate a certain fragment of DNA. It does not shed light on the synthetic activity of the polymerase pulled down. This is particularly relevant as polymerase epsilon has been characterized as existing in

two distinct conformations – one which is suggested to reflect an actively engaged and synthesizing catalytic domain, the other reflecting continued interaction with the CMG but disengagement from DNA [127]. These two conformations threaten the assumption that a switch in the polymerase carrying out replicative enzymatic function would be apparent in localization data from ChIP-Seq, as there is the potential for a polymerase epsilon complex to disengage from synthesis while remaining closely tethered to the area.

Second, capturing polymerase localization offers a snapshot in time. This can therefore shed light on which tagged polymerase is at a specific chromatin location at a given timepoint, however it cannot yield insight into where the given polymerase was before or how long it may have been on chromatin.

One especially compelling polymerase switching question is: where does switching occur? Are particular genomic contexts more prone to polymerase switching? These questions, unfortunately, are intractable given the described experimental system, as very tight timepoint sampling would be required to gain high resolution peaks and to have confidence that a peak represented the precise point of polymerase switching onto the chromatin.

Both of these weaknesses could be addressed by further experimentation using a complementary next generation sequencing approach, polymerase usage sequencing (Pu-Seq) [123]. Briefly, the Pu-Seq approach leverages polymerase mutations known to increase the propensity for the promiscuous incorporation of rNTPs in a ribonucleotide excision repair defective background. A sequencing library can then be constructed to map locations of rNTPs in the genome, and the identity (mutant or wildtype) of the enzyme responsible for the synthesis of a given

region of DNA can be inferred from the frequency of ribonucleotides in that region. Critically, this approach interrogates polymerase function rather than polymerase localization, alleviating the concerns of the first weakness. Additionally, by synthesizing cells in the absence of a mutant polymerase, releasing them into S phase, inducing the mutant polymerase, and capturing DNA at the end of replication, it would be possible to map tracts of polymerase usage. For a single cell, rNTP abundance would begin at the point of switching, and thus in looking at ensemble data genomic sites showing inflection points in rNTP abundance could indicate points of polymerase switching.

The approach as designed, or an approach as described utilizing Pu-Seq, will systematically underestimate switching behavior as a polymerase could be replaced by another polymerase with the same tag. To minimize this possibility, it would be possible to employ a transcriptional flip, where the expression of the initial 'endogenous' polymerase is turned off when the induced polymerase is turned on. This would increase the chances of a polymerase switch incorporating the newly induced polymerase by stopping the continued production of the endogenous polymerase. Alternatively, *in vivo* single molecule microscopy approaches – analogous to those undertaken in bacterial systems – could be used to assay for switching by determining the residence times of a specific polymerase complex on the chromatin [211]. In these approaches, the entire polymerase pool is tagged to allow for conjugation of a fluorescent molecule. The fluorescent molecule can then be introduced into the cell and excited in a manner to allow limited activation, such that the density of activated fluorescent tags is low enough to resolve and track single polymerase molecules over time in live cells. This excitingly allows for single-cell and single-molecule information but comes

with the limitation of yielding no information on genomic locus. This necessitates careful experimental design to identify specifically switching events, instead of replisome termination and initiation events *in vivo*.

### *Switching to other polymerases*

For a complete picture of polymerase switching, it will be important to consider switching from polymerase epsilon to other DNA polymerases. There is evidence to suggest that polymerase delta plays a role in initiation and termination on the leading strand [125], as well as in homologous recombination and potentially break induced replication [212]. Furthermore, a variety of translesion polymerases become important in replicating over damaged template [213]. This same approach could be taken with tagged or mutator versions of alternative polymerases to develop a full picture of the genome's synthetic landscape.

# Chapter V. Discussion

This work has aimed to push forward understanding of two regulatory mechanisms functioning in eukaryotic DNA replication to preserve genome integrity while allowing for efficient genome duplication.

First, the hypothesis that cohesin-mediated TAD architecture acts as a fundamental unit of organization for replication initiation, thereby driving replication timing across the genome, was tested [52]. Using an auxin-inducible cohesin degradation system, replication timing patterns were assessed with high-throughput sequencing and microscopy techniques following acute depletion of cohesin during ongoing replication and cohesin depletion from early G1. Replication timing patterns did not significantly change in the absence of cohesin, supporting the conclusion that cohesin-mediated genome architecture does not drive replication timing patterns genome-wide. This is in keeping with other independently-published findings [116,214] showing DNA replication timing patterns to be robust to genome-wide conformational changes driven by the loss of cohesin or CTCF. This leaves an open question of what mechanisms drive replication timing; compartmentalization remains a very compelling hypothesis. Identification of the genome characteristics and physical forces driving replication timing will be important not only to deepen understanding of how replication initiation is regulated and how limiting initiation factors are utilized, but also to unpick the relationship between replication timing and replicative errors such as copy number variation.

While copying template DNA, replication machinery must contend with a variety of insults and obstacles, from stressed conditions nucleus-wide to specific lesions on the template DNA. One method to mitigate the detriment of these potential

challenges is to employ an arsenal of replicative polymerases that allow for flexible adaptation to replication demands. The concept of switching synthesizing polymerase during ongoing replication is well-established in lagging strand synthesis and in bacterial systems, but the full extent of polymerase switching *in vivo* in eukaryotic leading strand synthesis has not been investigated. Thus, this work developed a polymerase induction system to allow for the direct investigation of leading strand polymerase switching *in vivo.* Pilot experiments using this system have echoed *in vitro* findings that the leading strand polymerase epsilon is stably associated with the CMG, as the results were consistent with a lack of widespread and rapid switching of polymerase epsilon during ongoing replication. This system is primed for further investigations to assess polymerase stability in the context of perturbations such as UV-induced template lesions.

Through investigating two aspects of replication regulation, this work attempts to advance understanding of the careful balance struck between preserving genome integrity and enabling efficient cell cycle progression. These two investigations highlight the robustness of the replicative program, and its adaptability to the various potential challenges to genome integrity that might be encountered in the course of DNA replication.

# Appendix

**Table 2. Reagents and buffers.**

| Buffer Name | Components | Protocol |
|---|---|---|
| Lysis buffer (smash & grab) | - 2% (v/v) Triton X-100<br>- 10 mM Tris pH 8.0<br>- 100 mM NaCl<br>- 1 mM EDTA<br>- 1% w/v SDS | Smash & Grab |
| Tris-EDTA, TE | - 10 mM Tris pH 8.0<br>- 1 mM EDTA | Smash & Grab<br>Repli-Seq |
| Wash buffer A | - 1 mM PMSF<br>- ddH2O<br>- 0.1% sodium azide<br>- 1 mM sodium vanadate | Chromatin fractionation |
| Spheroplasting buffer | - 50 mM potassium phosphate buffer (pH 7.5)<br>- 0.6 M sorbitol<br>- 10 mM DTT<br>- 1x complete ULTRA mini EDTA free<br>- 0.4 mg zymolyase 20-T / 100 mg wet cell pellet | Chromatin fractionation |
| 2x EB | - 100 mM HEPES KOH (pH 7.5)<br>- 200 mM KCl<br>- 5 mM $MgCl_2$ | Chromatin fractionation |
| 10x Running Buffer | - 1.92 M glycine<br>- 0.25 M tris<br>- 1% SDS | Western blot |
| Transfer Buffer | - 1x running buffer<br>- 20% methanol | Western blot |
| Acrylamide Gel : Stacking (6%) | - 1.11 mL $ddH_2O$<br>- 0.4 mL 30% acrylamide<br>- 0.5 M Tris (pH 6.8)<br>- 10 µL 10% SDS<br>- 20 µL 10% APS<br>- 2 µL TEMED | Western blot |
| Acrylamide Gel : Running (7%) | - 5 mL $ddH_2O$<br>- 2.34 30% acrylamide<br>- 2.5 mL 1.5 M Tris (pH 8.8)<br>- 50 µL 20% SDS<br>- 100 µL 10% APS<br>- 10 µL TEMED | Western blot |
| Dissection Buffer | - 0.8 M sorbitol | Tetrad dissection |

| | | |
|---|---|---|
| | - 10 mM EDTA<br>- 10 mM NaPO4 (pH 7.2) | |
| SPZ Buffer | - 2.5 mg/mL Zymolyase 100T<br>- 30 mL 2 M sorbitol<br>- 4.05 mL 1 M Na2HPO4<br>- 0.95 mL 1 M Na2H2PO4<br>Made up to 50 mL with $H_2O$ (pH 7.5) | Quick SPZ yeast DNA extraction |
| 2 x EB | - 100 mM HEPES KOH (pH 7.5)<br>- 200 mM KCl<br>- 5 mM $MgCl_2$<br>Made up in water. Add cOmplete protease inhibitor cocktail to 1xEB before use. | Chromatin fractionation |
| Wash buffer A | - 0.1% sodium azide<br>- 1 mM PMSF<br>Made up in water | Chromatin fractionation |
| Spheroplasting buffer | - 50 mM potassium phosphate<br>- 0.6 M sorbitol<br>- 10 mM DTT<br>- 0.4 mg / 100 mg cell pellet Zymolyase 20T<br>Made up in water | Chromatin fractionation |
| Sorbitol / EB | - 0.4 M sorbitol<br>- 1 mM PMSF<br>Made up in 1x EB | Chromatin fractionation |
| Buffer EB | - 1 mM PMSF<br>Made up in 1x EB | Chromatin fractionation |
| Sucrose Cushion<br>30% Sucrose / EB | - 30% sucrose<br>- 1 mM PMSF<br>Made up in 1x EB | Chromatin fractionation |
| EBX | - 0.25% Triton X-100<br>Made up in 1x EB | Chromatin fractionation |
| 4x LDS loading buffer | - 2 mL 1 M Tris HCl (pH 6.8)<br>- 0.8 g SDS<br>- 4.0 mL 100% glycerol<br>- 0.4 mL 14.7 M beta mercaptoethanol<br>- 1.0 mL 0.5 M EDTA<br>- 8 mg bromophenol blue<br>Made up to 10 mL with water | Chromatin fractionation |
| TBS | - 20 mM Tris-HCl (pH 7.4)<br>- 150 mM NaCl<br>Made up in water | ChIP-Seq |

| | | |
|---|---|---|
| Lysis Buffer (ChIP-Seq) | - 50 mM HEPES KOH (pH 7.5)<br>- 140 mM NaCl<br>- 0.1% Na-deoxycholate<br>- 1 mM EDTA<br>- 1% Triton X-100<br>- 1 mM PMSF<br>- Protease inhibitor cocktail<br>Made up in water | ChIP-Seq |
| PBS/BSA | 0.5% (w/v) BSA in 1x PBS | ChIP-Seq |
| Wash Buffer | - 10 mM Tris-HCl (pH 8.0)<br>- 250 mM LiCl<br>- 0.5% Na-deoxycholate<br>- 1 mM EDTA<br>Made up in water | ChIP-Seq |
| TE-SDS | - 10 mM Tris HCl (pH 8)<br>- 1 mM EDTA<br>- 1% SDS<br>Made up in water | ChIP-Seq |
| Dissection Buffer | - 0.8 M Sorbitol<br>- 10 mM EDTA<br>- 10 mM $NaPO_4$ (pH 7.2) | Tetrad dissection |
| 10x TBE | - 108 g/L Trizma base<br>- 55 g/L boric acid<br>- 50 mL/L 0.5 M EDTA (pH 8) | Agarose gel electrophoresis |
| Orange Loading Dye | - 40 % (w/v) glycerol<br>- 0.5 % SDS<br>- 10 mM EDTA<br>- 0.1 % (w/v) Orange G Sodium Salt | Agarose gel electrophoresis |
| Wash Solution | - 1% BSA<br>- 0.5% (v/v) Tween 20<br>- PBS | Mammalian Flow Cytometry |
| IP Buffer | - PBS<br>- 0.0625 % (v/v) Triton X-100 | Repli-Seq |
| Ponceau | - 0.1% (w/v) Ponceau Red powder<br>- 5% *(v/v)* acetic acid | Western blot |
| PBST | 0.1% (v/v) Tween 20 in PBS | Western blot |
| Blocking Buffer (Western) | 5% (w/v) milk in PBST | Western blot |
| Dissection Buffer | - 0.8 M sorbitol<br>- 10 mM EDTA<br>- 10 mM $NaPO_4$ (pH 7.2) | Tetrad dissection |

| | | |
|---|---|---|
| Sample Buffer | - 0.06 M Tris-HCl (pH 6.8)<br>- 5% glycerol<br>- 2% SDS<br>- 4% β-mercaptoethanol<br>- 0.0025% bromophenol blue | Yeast total protein extraction |
| 2x Digestion Buffer | - 20mM Tris HCl pH 8.0<br>- 0.2M EDTA<br>- 1% SDS<br>- 100μg/ml DNase free RNase A | Mammalian DNA extraction |
| IF Block Solution | 5% (*w/v*) BSA in PBST | Immunofluorescence |

**Table 3. Plasmid table.**

| Final Plasmid *isolate used | Description | Used for Strains | Construction Strategy (Template:F/R primers) |
|---|---|---|---|
| pPO01 | pROS16 double guide plasmid targeting the *S. cerevisiae HO* locus. His3 selection. | POY01, POY18, POY24, POY25, POY26, POY27, POY32, POY33 | Gibson cloning of pROS16 backbone with primer sets AC0060/AC0061, AC0060/AC0062. |
| pPO02*, pPO03 | Repair template to create *S. cerevisiae* genotype: *HO::pPOL2 POL2-E-Flag* | POY01, POY18 | Gibson cloning of parts: ACS001: PO45/46, ACS009: EH248/CN127, ACS009: digest BsiWI, NgoMIV |
| pPO05 | Repair template to create *S. cerevisiae* genotype: *HO::$P_{GalL}$ POL2-Flag* | NA | BsaI digest and T4 ligation of parts: pYM-N26: PO80/PO81, pEDH16, pYTK060, pYTK066, pYTK089, pPO02: PO88/PO89, pPO02: PO86/PO87 |
| pPO06 | Repair template to create *S. cerevisiae* genotype: *HO::$P_{GalS}$ POL2-Flag* | POY24, POY25 | BsaI digest and T4 ligation of parts: pYM-N30: PO82/PO83, pEDH16, pYTK060, pYTK066, pYTK089, pPO02: PO88/PO89, pPO02: PO86/PO87 |
| pPO07 | Repair template to create *S. cerevisiae* genotype: *HO::$P_{Gal1}$ POL2-Flag* | POY26, POY27 | BsaI digest and T4 ligation of parts: pYTK030: PO84/PO85, pEDH16, pYTK060, pYTK066, pYTK089, pPO02: PO88/PO89, pPO02: PO86/PO87 |

| | | | |
|---|---|---|---|
| pPO08/pPO09* | 6xFlag-9xMyc-KanMX | POY39 | Gibson cloning of parts:<br>PYM18 (9Myc): PO99/PO100,<br>6xFlag:PO101/PO102,<br>6xFlag digested with SalI/BamHI (gel extract 4kb fragment)<br><br>Screened with EH251/CN2198.<br>Digest screen with BamHI-HF.<br><br>Amplified with CA596/597 to confer homology to *S. bayanus MCM4*.<br>Transformed into JRY8149 under G418 selection to tag *S. bayanus MCM4* with Flag/Myc on the 3' end. |

**Table 4. Primer table.**

| Oligo Name | Part / Locus | Description | Sequence |
|---|---|---|---|
| PO11 | *DBF4* | *DBF4* 5’ end amplification F1 | GCCAACTACCCAAGGACACA |
| PO12 | *DBF4* | *DBF4* 5’ end amplification R1 | TGTTGAAACCAGCAATGGGG |
| PO13 | *DBF4* | *DBF4* 5’ end amplification F2 | TGACTTCCTTGAGCTACCAG G |
| PO14 | *DBF4* | *DBF4* 5’ end amplification R2 | CGAGCGCGTCTGACTTTAAA |
| PO15 | Gibson DBF4-HaloTag | Gibson - anneal *DBF4* 5' (F), overlap backbone | TTGTAAAACGACGGCCAGTG CGCCCCCCGGGGCCGCCC |
| PO16 | Gibson DBF4-HaloTag | Gibson - anneal *DBF4* 5' (R), overlap HaloTag | GCTTGGCCATGGCAGTCGCC CGGCACCTACATGCTGGGTC |
| PO17 | Gibson DBF4-HaloTag | GIBSON - anneal HaloTag (F), overlap *DBF4* 5' | GGCGACTGCCATGGCCAAG CCTTTGTCTC |
| PO18 | Gibson DBF4-HaloTag | Gibson - anneal HaloTag (R), overlap *DBF4* 3' | CGGAGTTCATGCCGGAAATC TCGAGCGTC |
| PO19 | Gibson DBF4-HaloTag | Gibson - silent mutation sgRNA1 anneal *DBF4* 3' (F), overlap HaloTag | GATTTCCGGCATGAACTGCG GAGCCATGAG |
| PO20 | Gibson DBF4-HaloTag | Gibson - anneal *DBF4* 3' (R), overlap backbone | CTATGACCATGATTACGCCA AGGATGCAGTCCGTCGCC |
| PO43 | pRos16 | pRos16 guide screening | GCCTAAGAATGTTCAAC |
| PO44 | pRos16 | pRos16 guide screening 2 | CTATCCGTCTACGTGAATA |
| PO45 | ACS001 (E tag) | ACS001 Gibson fragment F | CTGGTGCCGGTTCTGCTG |
| PO46 | ACS001 (E tag) | ACS001 Gibson fragment R | CTTGTAGTCTCCGTGATGAT GATGGTGGTGGGATCCCGTA CGAAGCTTCAG |
| PO47 | pYM18 (*POL2*) | pYM18 amplification *POL2* F | GGTTTTGATATATTATTGAGT TGTATTGCTGATTTGACCATA CGTACGCTGCAGGTCGAC |
| PO48 | pYM18 (*POL2*) | pYM18 amplification *POL2* R | TAAAGAGGCCATTGAACCTC GCGTTATATACTGCTTACTCA ATCGATGAATTCGAGCTCG |

| | | | |
|---|---|---|---|
| PO49 | *HO* locus | sequencing primer *HO* locus | CACGAGTAAGCGGCACTTG |
| PO50 | pMEL (*POL1*) | F guide pMEL series, *POL1* Kunkel mutation | GCAGTGAAAGATAAATGATC TTTGTATCCATCTATTATCCG TTTTAGAGCTAGAAATAGC |
| PO51 | pMEL (*POL1*) | R guide pMEL series, *POL1* Kunkel mutation | GCTATTTCTAGCTCTAAAACG GATAATAGATGGATACAAAG ATCATTTATCTTTCACTGC |
| PO52 | pMEL | pMEL plasmid amplification F | GATCATTTATCTTTCACTGCG GAGAAG |
| PO53 | pMEL | pMEL plasmid amplification R | GTTTTAGAGCTAGAAATAGC AAGTTAAAATAAGGCTAGTC |
| PO54 | *POL1* | *POL1* Kunkel mutation repair template amplification F | GCGGTACAAGAGCTGGTAGA |
| PO55 | *POL1* | *POL1* Kunkel mutation repair template amplification R | AGCCTTTTAAATCCCAAGCC A |
| PO56 | *POL1* guide | *POL1* guide for DB17 F | TGATCTTTGTATCCATCTATT ATCCG |
| PO57 | *POL1* guide | *POL1* guide for DB17 R | AAAACGGATAATAGATGGAT ACAAAG |
| PO58 | *POL2-Myc* | screening *POL2* 5' end for Myc tag incorporation | CAAGATTTGAGATGCTCCAG ATG |
| PO59 | *POL2-Myc* | screening *POL2* 5' end for Myc tag incorporation | ACCTCGCGTTATATACTGCTT AC |
| PO60 | *POL2* guide | guide for *POL2* 5' end, designed for pDB17 ligation | TGATCTTATATACTGCTTACT CATAG |
| PO61 | *POL2* guide | guide for *POL2* 5' end, designed for pDB17 ligation | AAAACTATGAGTAAGCAGTAT ATAAG |
| PO62 | *POL2* | *POL2* endogenous 3' end F | CCCAGAGAAAGCATTGTTCA AAA |
| PO62 | *POL2* | *POL2* endogenous 3' end F. Used as a sequencing primer to confirm pPO05, pPO06, pPO07. | CCCAGAGAAAGCATTGTTCA AAA |

| | | | |
|---|---|---|---|
| PO63 | *POL2* | *POL2* endogenous 3' end R | CCCTAAAAGCAACTTCCGGAG |
| PO63 | *POL2* | *POL2* endogenous 3' end R | CCCTAAAAGCAACTTCCGGAG |
| PO64 | $P_{GAL1}$ | R primer for $P_{GAL1}$ promoter amplification from pYTK30. Long overhang for joining to pP002 cut with PvuII. | GTTGCCAGCTGAATATCTTGCAGTGGAAGATCCTCCGTTGTTTTTTTTCTTGCCAAACATCATTATAGTTTTTTCTCCTTGACGTTAAAG |
| PO65 | $P_{GAL1}$ | F primer for $P_{GAL1}$ amplification from pYTK30. Long overhang for joining to left hand *HO* homology arm by fusion PCR. | GCCTTTGGACTTAAAATGGCGTCCCCATTATCTTAGCCTAAAAAAAC |
| PO66 | *HO* locus | R primer for *HO* homology arm amplification from pPO02. Long overhang for joining to $P_{GAL1}$ by fusion PCR. | GTTTTTTTAGGCTAAGATAATGGGGACGCCATTTTAAGTCCAAAGGC |
| PO67 | *HO* locus | F primer for *HO* homology arm amplification from pPO02. covers PacI site on 5' end to allow for fusion PCR product to be cut and ligated to PO02. | GCCGTTAATTAAAATTATCCTGGGC |
| PO68 | $P_{GalL}$ | F primer to amplify $P_{Gal1}$with type 1 5' overhang. | GGTCTCACCCTGAGCTCGTGACAGCCCTCCG |
| PO69 | $P_{GalL}$ | R primer to amplify $P_{GalL}$ with type 2 3' overhang. | GGTCTCACATAAATCCGGGGTTTTTTCTCCTTGACG |
| PO70 | $P_{GalS}$ | F primer to amplify $P_{GalS}$ with type 1 5' overhang. | GGTCTCACCCTCGGAAGACTCTCCTCCGTGCG |

| | | | |
|---|---|---|---|
| PO71 | $P_{GalS}$ | R primer to amplify *$P_{GalS}$* with type 2 3' overhang. | GGTCTCACATAGTTTTTTCTCCTTGACGTTAAAGTATAGAGG |
| PO72 | $P_{Gal1}$ | F primer to amplify *$P_{Gal1}$* with type 1 5' overhang. | GGTCTCACCCTCCCCATTATCTTAGCCTAAAAAAACC |
| PO73 | $P_{Gal1}$ | R primer to amplify *$P_{Gal1}$* with type 2 3' overhang. | GGTCTCACATATATAGTTTTTTCTCCTTGACGTTAAAG |
| PO74 | *HO* locus | F primer to amplify MoClo's *HO* 3' homology arm, subbing in Type 5 overhang. | GGTCTCAGCTGTATCGTGTTGCATCTGCGGC |
| PO75 | *HO* locus | R primer to amplify MoClo's *HO* 3' homology arm with existing type 7 overhang. | GGTCTCATCGGGGCAGAACTAAC |
| PO76 | *HO* locus | F primer to amplify 5' *HO* homology arm from existing BsaI cut site with Type 8b 5' overhang (CAAT) | GGTCTCACAATAACACGTAAGGTTAAGATGTC |
| PO77 | *HO* locus | R primer to amplify 5' *HO* homology arm adding BsaI cut site with Type 8b 3' overhang (GGGA) | GGTCTCAAGGGACGCCATTTTAAGTCCAAAGGCAC |
| PO78 | *HO* locus | F primer to amplify 3' *HO* homology arm from pPO02 adding BsaI cut site with Type 5 5' overhang | GGTCTCAGCTGCTGGGGGAACAACTTCACAGAATG |
| PO79 | *HO* locus | R primer to amplify 3' *HO* homology arm from pPO02 adding BsaI cut site with Type 7 3' overhang | GGTCTCATCGGCTGTAAGATTCCGCCACATTTTATAC |

| | | | |
|---|---|---|---|
| PO80 | $P_{GalL}$ | F primer to amplify $P_{GalL}$ with type 1 5' overhang - BsmbI cut site to create parts vector. | GCATCGTCTCAGGTCTCACC CTGAGCTCGTGACAGCCCTC CG |
| PO81 | $P_{GalL}$ | R primer to amplify $P_{GalL}$ with type 2 3' overhang - BsmbI cut site to create parts vector. | ATGCCGTCTCAGGTCTCACA TAAATCCGGGGTTTTTTCTCC TTGACG |
| PO82 | $P_{GalS}$ | F primer to amplify $P_{GalS}$ with type 1 5' overhang - BsmbI cut site to create parts vector. | ATGCCGTCTCAGGTCTCACC CTCGGAAGACTCTCCTCCGT GCG |
| PO83 | $P_{GalS}$ | R primer to amplify $P_{GalS}$ with type 2 3' overhang – BsmbI cut site to create parts vector. | ATGCCGTCTCAGGTCTCACA TAGTTTTTTCTCCTTGACGTT AAAGTATAGAGG |
| PO84 | $P_{Gal1}$ | F primer to amplify $P_{Gal1}$ with type 1 5' overhang - BSMB1 cut site to create parts vector. | ATGCCGTCTCAGGTCTCACC CTCCCCATTATCTTAGCCTAA AAAAACC |
| PO85 | $P_{Gal1}$ | R primer to amplify $P_{Gal1}$ with type 2 3' overhang - BsmbI cut site to create parts vector. | ATGCCGTCTCAGGTCTCACA TATATAGTTTTTTCTCCTTGA CGTTAAAG |
| PO86 | *HO* locus homology arms | F primer to amplify 3' *HO* homology arm from pPO02 adding BsaI cut site with Type 5 5' overhang - Bsmb1 cut site to create parts vector. | ATGCCGTCTCAGGTCTCAGC TGCTGGGGGAACAACTTCAC AGAATG |
| PO87 | *HO* locus homology arms | R primer to amplify 3' *HO* homology arm from pPO02 adding BsaI cut | ATGCCGTCTCAGGTCTCATC GGCTGTAAGATTCCGCCACA TTTTATAC |

| | | | |
|---|---|---|---|
| | | site with Type 7 3' overhang - Bsmb1 cut site to create parts vector. | |
| PO88 | *HO* locus homology arms | F primer to amplify 5' *HO* homology arm from existing BsaI cut site with Type 8b 5' overhang - Bsmb1 cut site to create parts vector. | ATGCCGTCTCAGGTCTCACA ATAACACGTAAGGTTAAGAT GTC |
| PO89 | *HO* locus homology arms | R primer to amplify 5' *HO* homology arm adding BsaI cut site with Type 8b 3' overhang - BSMB1 cut site to create parts vector. | ATGCCGTCTCAGGTCTCAAG GGACGCCATTTTAAGTCCAA AGGCAC |
| PO90 | $P_{GalL}$ | F primer to amplify $P_{GalL}$ with overhang to fit Carolin's *HO* vector + BsmbI cut site. Modification of PO80 | GCATCGTCTCAGGTCTCACT GAGAGCTCGTGACAGCCCTC CG |
| PO91 | $P_{GalL}$ | F primer to amplify $P_{GalL}$ with overhang to fit Carolin's *HO* vector + BsmbI cut site. Modification of PO82 | ATGCCGTCTCAGGTCTCACT GACGGAAGACTCTCCTCCGT GCG |
| PO92 | $P_{GalL}$ | F primer to amplify $P_{GalL}$ with overhang to fit Carolin's *HO* vector + BsmbI cut site. Modification of PO84 | ATGCCGTCTCAGGTCTCACT GACCCCATTATCTTAGCCTAA AAAAACC |
| PO93 | tTDH1 | F primer to amplify terminator (pYTK066, tTDH1) with | CCGTAGTCGGTCTCCTGGCA TAAAG |

| | | | |
|---|---|---|---|
| | | BsmbI and BsaI cut site and type 4b overhangs | |
| PO94 | tTDH1 | R primer to amplify terminator (pYTK066, *tTDH1*) with BsmbI and BsaI cut site and overhang for Carolin's *HO* vector | GTTTTTTATTGGTCTGGTCTCATGCTCGTTCAGGGTAATATATTTTAAC |
| PO95 | pYM18 (*MCM4-Flag-Myc*) | F primer for amplification of pYM18 to add Myc tag to *MCM4-Flag* in CAY300 | GACTATAAGGACGACGACGACAAGTATTCCGGTTCTGCTGCTAGTGGTGAAC |
| PO96 | pYM18 (*MCM4-Flag-Myc*) | R primer to for amplification of pYM18 to add Myc tag to *MCM4-Flag* in CAY300. | TACATTATACGAAGTTATATTAAGGGTTGTCGACTCACCACTTTTTAGCTAGTGGATCCG |
| PO97 | pMEL16 (*MCM4-Flag* guide) | sgRNA (F) for CAY300 *MCM4-Flag* - construct into pMEL16 using https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4399441/ | GCAGTGAAAGATAAATGATCCACAATTTTAAAGTTTAAAAGTTTTAGAGCTAGAAATAGC |
| PO98 | pMEL16 (*MCM4-Flag* guide) | sgRNA (R) for CAY300 *MCM4-Flag* - construct into pMEL16 using https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4399441/ | GCTATTTCTAGCTCTAAAACTTTTAAACTTTAAAATTGTGGATCATTTATCTTTCACTGC |
| PO99 | pYM18 (9Myc) | Gibson 6Flag-9Myc: pYM18 F primer | AAGGACTATAAGGACGACGACGACAAGTCCGGTTCTGCTGCTAGTGGTG |
| PO100 | pYM18 (9Myc) | Gibson 6Flag-9Myc: pYM18 R primer | ATAGCATACATTATACGAAGTTATATTAAGGGTTGTTTAGCTAGTGGATCCGTTCAAGTC |
| PO101 | 6xFlag | Gibson 6Flag-9Myc: 6Flag F primer | GCTGAAGCTTCGTACGGGATCCCACC |

| | | | |
|---|---|---|---|
| PO102 | 6xFlag | Gibson 6Flag-9Myc: 6Flag R primer | CCACTAGCAGCAGAACCGGACTTGTCGTCGTCGTCCTTATAGTCCTT |
| PO103 | 6xFlag (*POL2*) | Flag tag endogenous *POL2* (amplify 6xFlag with HA for endogenous *POL2*) | TTGATATATTATTGAGTTGTATTGCTGATTTGACCATACACCACCATCATCATCACGGAG |
| PO104 | 6xFlag (*POL2*) | Flag tag endogenous pol2 (amplify 6xFlag with HA for endogenous *POL2*) | AGAGGCCATTGAACCTCGCGTTATATACTGCTTACTCAACTATAGGGAGACCGGCAGATC |
| PO105 | *GAL2* | *GAL2* upstream F primer (for Gal2 deletion) | TAAACACAAGATTAACATAATAAAAAAAATAATTCTTTCATAATG |
| PO106 | *GAL2* | *GAL2* downstream R primer (for Gal2 deletion) | GAAAAATTAAGAGAGATGATGGAGCGTCTCACTTCAAACGCATTA |
| PO107 | *MCM4* guide | *S. bayanus* WT *MCM4* 3' guide F | GCAGTGAAAGATAAATGATCGTATCAGAAAGTATCAGACAGTTTTAGAGCTAGAAATAGC |
| PO108 | *MCM4* guide | *S. bayanus* WT MCM4 3' guide R | GCTATTTCTAGCTCTAAAACTGTCTGATACTTTCTGATACGATCATTTATCTTTCACTGC |
| PO109 | 9xMyc (pYM18) | Amplify 9xMyc (pYM18) with homology to JRY8149 *MCM4* 3' (F) | TGTTCTTGGTGAAGGTGTAAGGAGATCTGTTCGCTTAAACAACCGTGTCTCCGGTTCTGC |
| PO110 | 9xMyc (pYM18) | Amplify 9xMyc (pYM18) with homology to JRY8149 *MCM4* 3' (R) | TATTAAAAATGTTACAAAGCACACAAATGTGAGTTATACAGGCAATCGATGAATTCGAGC |
| PO111 | NatMX | Screening *NatMX* marker swap | CCTCGGCGTCCCCCGGG |
| PO112 | NatMX | Screening *NatMX* marker swap | TGGCATTATACTCCTGCTAGAAAGT |
| PO113 | *MCM4* | Screening pMel *MCM4 S. bayanus* | TTTTAAACTTTAAAATTGTG |
| PO114 | *MCM4* | Screening pMel *MCM4 S. bayanus* | GTATCAGAAAGTATCAGACAGTTTTAGAGC |

| | | | |
|---|---|---|---|
| PO115 | *GAL2* | *GAL2* confirmation primer B | AAAGATACCTGCAGTAATCATCAGC |
| PO116 | *GAL2* | *GAL2* confirmation primer D | TCTGGTGATGTGGTCCTTTAATAAT |
| PO117 | *MCM4* | Screening pMel MCM4 *S. bayanus* (guide specific, R) | TGTCTGATACTTTCTGATAC |
| PO118 | *MCM4* | Screening pMel *MCM4 S. bayanus* (nonspecific, R) | CGTCCCAAAACCTTCTCAAGC |
| EH248 | *E-tag* | used with CN127 to amplify ACS009 backbone for the creation of the pPO02/pPO03 plasmid (*HO::pPOL2 pol2-E tag-Flag*) | AGGCTGTTCTTCTTAGAGCG |
| EH251 | *pSP6* | used with CN2198 to confirm creation of pPO08, pPO09 | GTGACACTATAGAACGCGGC |
| EH287 | *POL2* | Used as a sequencing primer to confirm pPO05, pPO06, pPO07. | GCATCGTCTCAAGATACTAGCGTTGCGTGCATGTTCG |
| DB85 | *ars607, ARS607* | *ars607* (used with AC39), *ARS607* (used with AC40) | TTTAGCTGGGTTTATGGGAGG |
| CN127 | ACS009 | used with EH248 to amplify ACS009 backbone for the creation of the pPO02/pPO03 plasmid (HO::pPOL2 pol2-E tag-Flag) | CAGATGTCATAAAGTGAAACGTGAC |
| CN2031 | *ars606* | *ars606* (used with CN2036) | GAAATTGCAAACGAACTCGAG |
| CN2032 | *ARS606* | *ARS606* (used with CN2036) | gcgaaattgcaaacgaaaatata |
| CN2036 | *ars606, ARS606* | *ars606* (used with CN2031), *ARS606* (used with CN2032) | GCGCAGACGTATCAAACCTAA |

| | | | |
|---|---|---|---|
| CN2169 | pAG25 | Marker swap *NatMX* (used with CN2171 to amplify pAG25) | CAGCTGAAGCTTCGTACGC |
| CN2171 | pAG25 | Marker swap *NatMX* (used with CN2169 to amplify pAG25) | GCATAGGCCACTAGTGGATC TG |
| CN2198 | *KanMX* | used with EH251 to confirm creation of pPO08, pPO09 | GAGCCGTAATTTTTGCTTCG |
| CA26 | *ars608* | *ars608* (used with CA53) | GCGCATAGAAGAAAACTCGA G |
| CA53 | *ars608, ARS608* | *ars608* (used with CA26), *ARS608* (used with AC42) | TTGTTGTTGATTGGGTGGAT |
| CA377 | *MATa, MATα* | mating type a (used with CA379), mating type alpha (used with CA378) | AGTCACATCAAGATCGTTTAT GG |
| CA378 | *MATα* | mating type alpha (used with CA377) | GCACGGAATATGGGACTACT TCG |
| CA379 | *MATa* | mating type a (used with CA377) | ACTCCACTTCAAGTAAGAGTT TG |
| CA596 | *MCM4-6xFlag-9xMyc* | amplified pPO08/pPO09 with HA to *S. bayanus MCM4* and transformed amplicon to create spike-in strain POY39 | AGGCTGCAACAAGAGGATAA GGTTATTGTTCTTGGTGAAG GTGTAAGGAGATCTGTTCGC TTAAACAACCGTGTCTCCCA CCACCATCATCATCAC |
| CA597 | *MCM4-6xFlag-9xMyc* | amplified pPO08/pPO09 with HA to *S. bayanus MCM4* and transformed amplicon to create spike-in strain POY39 | AAAAAAGCCGCCCACCAAGT CAAGCTTCTAATATTAAAAAT GTTACAAAGCACACAAATGT GAGTTATACAGGCAACTATA GGGAGACCGGCAGATC |
| CA605 | *S. bayanus MCM4* 3’ | used with CA606 to confirm Flag-Myc tag on *S. bayanus* transformants | TCGGACTCAATGTCGTTCAA |

| | | | |
|---|---|---|---|
| | | POY37, POY38, POY39 | |
| CA606 | *S. bayanus MCM4 3'* | used with CA605 to confirm Flag-Myc tag on *S. bayanus* transformants POY37, POY38, POY39 | GGTTGGCTCGCAAAGTTATG |
| ALU17 | *RAD5* | *RAD5* F, used with ALU18 | GCAGCAGGACCATGTAAACG |
| ALU18 | *RAD5* | *RAD5* R, used with ALU17 | AAACTCGTTACTCCACTGCG |
| ALU42 | *BUD4* | *BUD4* F, used with ALU43 | CTTAATGGCGAAGATTCTGG |
| ALU43 | *BUD4* | *BUD4* R, used with ALU42 | GTCCTTGGTTCAGTTATACG |
| AC13 | *ARS609* | *ARS609* (used with AC15) | GAATTTAGCCAATTCCGGAAAAAACAT |
| AC14 | *ars609* | *ars609* (used with AC15) | GCCAATTCCGGAAAAACTCGAG |
| AC15 | *ars609, ARS609* | *ars609* (used with AC14), *ARS609* (used with AC13) | ACATACTTCTGTGCCCATAGC |
| AC27 | *POL2* | R primer annealing to the *POL2* gene. Used with PO49 to screen the pPO05/pPO06/pPO07 transformants for appropriate incorporation of the Gal(x)-Pol2-Flag construct at the *HO* locus. Used as a sequencing primer to confirm pPO05, pPO06, pPO07. | TTGAGAGTGTGTTGTACTTGTTGC |
| AC39 | *ars607* | *ars607* (used with DB85) | CTTGATCCCAACGTAACTAAATCTC |
| AC40 | *ARS607* | *ARS607* (used with DB85) | ACTTGATCCCAACGTAACTAAATATAAAC |
| AC42 | *ARS608* | *ARS608* (used with CA53) | CTTGCGCATAGAAGAAAACTAAAAGT |

| | | | |
|---|---|---|---|
| AC60 | pRos16 (*HO* locus) | used in creation of pRos16 guide plasmid pPO01 for the *HO* locus | GATCATTTATCTTTCACTGCGG |
| AC61 | pRos16 (*HO* locus) | used in creation of pRos16 guide plasmid pPO01 for the *HO* locus | GACGACCAGGTCAGCTAGGGGTTTTAGAGCTAGAAATAGCAAGTT |
| AC62 | pRos16 (*HO* locus) | used in creation of pRos16 guide plasmid pPO01 for the *HO* locus | CCTTTGGACTTAAAATGGCGGTTTTAGAGCTAGAAATAGCAAGTT |
| AC77 | *ars608.5, ARS608.5* | *ars608.5* (used with AC79), *ARS608.5* (used with AC78) | GAATTGATCGGTACCAGAGC |
| AC78 | *ARS608.5* | *ARS608.5* (used with AC77) | ATGTTTATTAGCTAGGATTATACACAT |
| AC79 | *ars608.5* | *ars608.5* (used with AC77) | TTATTAGCTAGGATTATACACTCGAG |

**Table 5. Yeast strain table.** In grey are yeast strains that were not generated in the course of this work but were important resources for the work carried out herein and thus are included for reference.

| Strain Name | Descriptive Name (relevant genotype) | Description | Parent |
|---|---|---|---|
| POY01/POY02 | *POL2-E-Flag*<br>*HO::POL2-E-Flag* | VGY129 (5 ori deletion) transformed with Cas9-Trp, pPO01, and pPO02 to introduce a Pol2 with an E tag and Flag tag at the C terminus at the HO locus.<br>Following transformation, yeast were plated onto drop-out plates supplemented with methionine and leucine.<br>Screened with PCR and Western blotting. | VGY129 |
| POY18/POY19 | *POL2-E-Flag;POL2-Myc*<br>*POL2-9xMyc-kanMX*<br>*HO::POL2-E-Flag* | Myc tag added to endogenous Pol2 on parent POY01 (*HO::Pol2-Flag-E*).<br>PYM18 (9-Myc tag followed by KAN resistance cassette) amplified with PO47/48 to confer homology arms to endogenous pol2 locus. | POY01 |
| POY20/POY21 | *POL2-Myc*<br>*POL2-9xMyc-kanMX* | Myc tag added to endogenous Pol2 on VGY129.<br>PYM18 (9-Myc tag followed by KAN resistance cassette) amplified | VGY129 |

| | | | |
|---|---|---|---|
| | | with PO47/48 to confer homology arms to endogenous pol2 locus. | |
| POY24/POY25 | *$P_{GalS}$-POL2-Flag;POL2-Myc*<br>*HO::$P_{GalS}$-POL2-Flag*<br>*POL2-9xMyc-kanMX* | GalS-pol2-Flag introduced at the *HO* locus into a parent strain with the endogenous Pol2 Myc tagged. POY20 was CRISPR transformed with pPO06 (galS-pol2-Flag inserted at the *HO* locus) using guide plasmid pPO01 and Cas9-Trp. SPZ confirmation (Primer Table: PO49/AC27) from *HO* locus outside of homology arms across the gal promoter to pol2. Further confirmed by Western blotting following 8 hours of galactose induction. | POY20 |
| POY26/POY27 | *$P_{Gal1}$-POL2-Flag:POL2-Myc*<br>*HO::$P_{GAL1}$-POL2-Flag*<br>*POL2-9xMyc-kanMX* | Gal1-pol2-Flag introduced at the *HO* locus into a parent strain with the endogenous Pol2 Myc tagged. POY20 was CRISPR transformed with pPO07 (gal1-pol2-Flag inserted at the *HO* locus) using guide plasmid pPO01 and Cas9-Trp. SPZ confirmation (Primer Table: PO49/AC27) from | POY20 |

| | | | |
|---|---|---|---|
| | | *HO* locus outside of homology arms across the gal promoter to pol2. Further confirmed by Western blotting following 8 hours of galactose induction. | |
| POY28/POY29 | Flag Norm *POL2-6xFlag-kanMX* | VGY129 transformed with 6xFlag-KanMX plasmid amplified with PO103/PO104. Selected with G418 and screened by PCR (Primer Table: CN127/CN2198). | VGY129 |
| POY30/POY31 | Gal2 NatMX knockout *gal2::NatMX* | S288C gal2::KanMX transformed with pAG25 amplified with CN2169/CN2171 to achieve a marker swap from KanMX to NatMX casettes for gal2 knockout. Selected with CloNAT, screened with PCR (Primer Table: PO111/PO112). | S288C gal2::KanMX |
| POY32/POY33 | *$P_{GalS}$-POL2-Flag;POL2-Myc;Δgal2* *POL2-Myc-KanMX* *HO::$P_{GalS}$-POL2-Flag* *gal2::NatMX* | POY25 transformed with PCR amplicon from *gal2::NatMX* strain (Primer Table: PO105/PO106) to knockout Gal2 with NatMX. Plated on G418 and Nat, to select for *pol2-9xMyc-KanMX* and *gal2::NatMX* | POY25 |

| | | | |
|---|---|---|---|
| | | genotypes, respectively. Screened by PCR (Primer Table: PO111/PO112). | |
| POY34/POY35/POY36 | *POL2-Myc;Δgal2*<br>*POL2-Myc-KanMX*<br>*gal2::NatMX* | POY20 transformed with PCR amplicon from *gal2::NatMX* strain (Primer Table: PO105/PO106) to knockout Gal2 with NatMX.<br>Plated on G418 and Nat, to select for *pol2-9xMyc-KanMX* and *gal2::NatMX* genotypes, respectively. Screened by PCR (Primer Table: PO111/PO112). | POY20 |
| POY37/POY38/POY39 | Flag-Myc Dual Spike-In<br>*MCM4-Flag-Myc* | JRY8149 transformed with pPO09 amplicon to confer homology arms to *S. bayanus MCM4* (Primer Table: CA596/597).<br>*S. bayanus* resuspended in transformation mix + PEG and incubated at 37°C for 40 min, plated onto YPAD and replica plated onto G418 ~12 hours later.<br>SPZ screening PCR across *MCM4* 3' (Primer Table: CA605/CA606).<br>Western of total protein extraction | JRY8149 |

| | | | |
|---|---|---|---|
| | | probed for Flag and Myc. | |
| POY40 | *cdc7-1; MATα; ars607::ars607 ACS->XhoI; ars608::ars608 ACS->XhoI; ars609::ars609 ACS->XhoI; ars610::URA3-ars610; BUD4; RAD5* | K2032 and ACY165 mated to get the *cdc7 ts* mutation into the ultra-long replicon background. Screened for: *cdc7 ts* (replica plating, growth defect at 30°C) *ars610-URA* (replica plating, growth on -URA) ars607 (Primer Table: DB85/AC39, DB85/AC40) ars608 (Primer Table: CA53/CA26, CA53/AC42) ars609 (Primer Table: AC13/AC15, AC14/AC15) rad5 (Primer Table: ALU17/ALU18, MnlI digest) bud4 (Primer Table: ALU42/ALU43, sequenced) MATα (Primer Table: CA377/CA379, CA377/CA378) | K2032 x ACY165 |
| POY41 | *cdc7-1; MATα; ars607::ars607 ACS->XhoI, ars608::ars608 ACS->XhoI, ars609::ars609 ACS->XhoI, ars610::URA3-ars610, bud4, RAD5* | K2032 and ACY165 mated to get the *cdc7 ts* mutation into the ultra-long replicon background. Screened for: *cdc7 ts* (replica plating, growth defect at 30°C) | K2032 x ACY165 |

| | | | |
|---|---|---|---|
| | | *ars610-URA* (replica plating, growth on -URA) ars607 (Primer Table: DB85/AC39, DB85/AC40) ars608 (Primer Table: CA53/CA26, CA53/AC42) ars609 (Primer Table: AC13/AC15, AC14/AC15) rad5 (Primer Table: ALU17/ALU18, MnlI digest) bud4 (Primer Table: ALU42/ALU43, sequenced) MATα (Primer Table: CA377/CA379, CA377/CA378) | |
| POY42 | *cdc7-1; MATα; RAD5; BUD4* | K2032 and ACY165 mated to get the *cdc7 ts* mutation into the ultra-long replicon background. Screened for: *cdc7 ts* (replica plating, growth defect at 30°C) *ars610-URA* (replica plating, growth on -URA) *ars607* (Primer Table: DB85/AC39, DB85/AC40) *ars608* (Primer Table: CA53/CA26, CA53/AC42) *ars609* (Primer Table: | K2032 x ACY165 |

| | | | |
|---|---|---|---|
| | | AC13/AC15, AC14/AC15) *rad5* (Primer Table: ALU17/ALU18, MnlI digest) *bud4* (Primer Table: ALU42/ALU43, sequenced) *MATα* (Primer Table: CA377/CA379, CA377/CA378) | |
| ACY165 | *MATα: ars607::ars607 ACS->XhoI ars608::ars608 ACS->XhoI ars609::ars609 ACS->XhoI, ars610::URA3-ars610 leu2, RAD5, BUD4* | | AM337 x ACY157 |
| VGY129 | 5 Ori Delta *MATa: ars607::ars607 ACS->XhoI ars608::ars608 ACS->XhoI ars608.5::ars608.5 ACS->XhoI ars609::ars609 ACS->XhoI, ars610::URA3-ars610, BUD4, RAD5* | ACY162 modified by CRISPR/Cpf1 system to inactivate ARS608.5 with XhoI site. * work done in Anne Donaldson lab | ACY162 |

# References


1. Keszthelyi A, Minchell N, Baxter J. The causes and consequences of topological stress during DNA replication. *Genes (Basel)*. 2016;7(12):134. doi:10.3390/genes7120134
2. Cortez D. Replication-coupled DNA repair. *Mol Cell*. 2019;74(5):866-876. doi:10.1016/j.molcel.2019.04.027
3. Klosin A, Reis K, Hidalgo-Carcedo C, Casas E, Vavouri T, Lehner B. Impaired DNA replication derepresses chromatin and generates a transgenerationally inherited epigenetic memory. *Sci Adv*. 2017;3(8). doi:10.1126/sciadv.1701143
4. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. *Nat Rev Genet*. 2009;10(8):551-564. doi:10.1038/nrg2593
5. Bell SP, Labib K. Chromosome duplication in *Saccharomyces cerevisiae*. *Genetics*. 2016;203(3):1027-1067. doi:10.1534/genetics.115.186452
6. MacAlpine DM, Almouzni G. Chromatin and DNA replication. *Cold Spring Harb Perspect Biol*. 2013;5(8). doi:10.1101/cshperspect.a010207
7. Guillou E, Ibarra A, Coulon V, et al. Cohesin organizes chromatin loops at DNA replication factories. *Genes Dev*. 2010;24(24):2812-2822. doi:10.1101/gad.608210
8. Yeeles JTP, Deegan TD, Janska A, Early A, Diffley JFX. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature*. 2015;519(7544):431-435. doi:10.1038/nature14285
9. Muramatsu S, Hirai K, Tak YS, Kamimura Y, Araki H. CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol ε, and GINS in budding yeast. *Genes Dev*. 2010;24(6):602-612. doi:10.1101/gad.1883410
10. Gambus A, Jones RC, Sanchez-Diaz A, et al. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol*. 2006;8(4):358-366. doi:10.1038/ncb1382
11. Sengupta S, Van Deursen F, De Piccoli G, Labib K. Dpb2 integrates the leading-strand DNA polymerase into the eukaryotic replisome. *Curr Biol*. 2013;23(7):543-552. doi:10.1016/j.cub.2013.02.011
12. Mantiero D, MacKenzie A, Donaldson A, Zegerman P. Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *EMBO J*. 2011;30(23):4805-4814. doi:10.1038/emboj.2011.404
13. Tanaka S, Nakato R, Katou Y, Shirahige K, Araki H. Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. *Curr Biol*. 2011;21(24):2055-2063. doi:10.1016/j.cub.2011.11.038
14. Rhind N, Gilbert DM. DNA replication timing. *Cold Spring Harb Perspect Med*. 2013;3(7):1-26. doi:10.1016/j.cell.2013.02.038
15. Ge XQ, Jackson DA, Blow JJ. Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev*. 2007;21(24):3331-3341. doi:10.1101/gad.457807
16. Okazaki R, Okazaki T, Sakabe K, Sugimoto K, Sugino A. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc Natl Acad Sci*. 1968;59(2):598-605. doi:10.1073/pnas.59.2.598

17. Johansson E, Dixon N. Replicative DNA polymerases. *Cold Spring Harb Perspect Biol*. 2013;5(6):a012799. doi:10.1101/cshperspect.a012799
18. Garbacz MA, Lujan SA, Burkholder AB, et al. Evidence that DNA polymerase δ contributes to initiating leading strand DNA replication in *Saccharomyces cerevisiae*. *Nat Commun*. 2018;9(1):858. doi:10.1038/s41467-018-03270-4
19. BioLabs NE. NEB Tm Calculator. https://tmcalculator.neb.com/#!/main. Published 2020. Accessed August 3, 2020.
20. Lee ME, DeLoache WC, Cervantes B, Dueber JE. A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth Biol*. 2015;4(9):975-986. doi:10.1021/sb500366v
21. Peace JM, Villwock SK, Zeytounian JL, Gan Y, Aparicio OM. Quantitative BrdU immunoprecipitation method demonstrates that Fkh1 and Fkh2 are rate-limiting activators of replication origins that reprogram replication timing in G1 phase. *Genome Res*. 2016;26(3):365-375. doi:10.1101/gr.196857.115
22. Schacherer J, Ruderfer DM, Gresham D, Dolinski K, Botstein D, Kruglyak L. Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. *PLoS One*. 2007;2(3):322. doi:10.1371/journal.pone.0000322
23. Gietz RD, Schiestl RH. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc*. 2007;2(1):35-37. doi:10.1038/nprot.2007.14
24. Kushnirov V V. Rapid and reliable protein extraction from yeast. *Yeast*. 2000;16(9):857-860. doi:10.1002/1097-0061(20000630)16:9<857::AID-YEA561>3.0.CO;2-B
25. Marchal C, Sasaki T, Vera D, et al. Genome-wide analysis of replication timing by next-generation sequencing with E/L Repli-seq. *Nat Protoc*. 2018;13(5):819-839. doi:10.1038/nprot.2017.148
26. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635
27. Karimzadeh M, Ernst C, Kundaje A, Hoffman MM. Umap and Bismap: quantifying genome and methylome mappability. *Nucleic Acids Res*. 2018;46(20):e120-e120. doi:10.1093/nar/gky677
28. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018.
29. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2016.
30. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996-1006. doi:10.1101/gr.229102
31. Dekker J, Belmont AS, Guttman M, et al. The 4D nucleome project. *Nature*. 2017;549(7671):219-226. doi:10.1038/nature23884
32. Consortium TEP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74. doi:10.1038/nature11247
33. Karolchik D, Hinrichs AS, Furey TS, et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*. 2004;32(Database issue):D493-6. doi:10.1093/nar/gkh103
34. Rao SSP, Huang S-C, Glenn St Hilaire B, et al. Cohesin loss eliminates all loop domains. *Cell*. 2017;171(2):305-320.e24. doi:10.1016/J.CELL.2017.09.026
35. Zerbino DR, Achuthan P, Akanni W, et al. Ensembl 2018. *Nucleic Acids*

*Res*. 2018;46(D1):D754-D761. doi:10.1093/nar/gkx1098
36. Neph S, Kuehn MS, Reynolds AP, et al. BEDOPS: High-performance genomic feature operations. *Bioinformatics*. 2012;28(14):1919-1920. doi:10.1093/bioinformatics/bts277
37. Wang Y, Song F, Zhang B, et al. The 3D Genome Browser: A web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol*. 2018;19(1):151. doi:10.1186/s13059-018-1519-9
38. Durand NC, Shamim MS, Machol I, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 2016;3(1):95-98. doi:10.1016/j.cels.2016.07.002
39. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359. doi:10.1038/nmeth.1923
40. Picard Toolkit. 2019. https://github.com/broadinstitute/picard. Accessed October 8, 2020.
41. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352
42. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842. doi:10.1093/bioinformatics/btq033
43. Hu B, Petela N, Kurze A, Chan K-L, Chapard C, Nasmyth K. Biological chromodynamics: A general method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic Acids Res*. 2015;43:132. doi:10.1093/nar/gkv670
44. Müller CA, Boemo MA, Spingardi P, et al. Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. *Nat Methods*. 2019;16(5):429-436. doi:10.1038/s41592-019-0394-y
45. Batrakou DG. CRAN - Package Repliscope. https://cran.r-project.org/web/packages/Repliscope/index.html. Published 2019. Accessed July 16, 2020.
46. Weddington N, Stuy A, Hiratani I, Ryba T, Yokochi T, Gilbert DM. ReplicationDomain: a visualization tool and comparative database for genome-wide replication timing data. *BMC Bioinformatics*. 2008;9:530. doi:10.1186/1471-2105-9-530
47. UCSC Genome Browser. CpG islands tracks. https://genome-euro.ucsc.edu/cgi-bin/hgTrackUi?hgsid=243985730_JfC5ND2GbsRqnTsFNahSuUYzzpaP&g=cpgIslandSuper. Accessed October 8, 2020.
48. UCSC Genome Browser. Schema for DNase HS - DNase I Hypersensitivity in 95 cell types from ENCODE. https://genome-euro.ucsc.edu/cgi-bin/hgTables?db=hg38&hgta_group=regulation&hgta_track=wgEncodeRegDnase&hgta_table=wgEncodeRegDnaseUwHct116Hotspot&hgta_doSchema=describe+table+schema. Accessed October 8, 2020.
49. Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155(4):934-947. doi:10.1016/J.CELL.2013.09.053
50. Boemo MA, Cardelli L, Nieduszynski CA. The Beacon Calculus: A formal method for the flexible and concise modelling of biological systems. Marucci L, ed. *PLOS Comput Biol*. 2020;16(3):e1007651. doi:10.1371/journal.pcbi.1007651

51. Hawkins M, Retkute R, Müller CA, et al. High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell Rep*. 2013;5(4):1132-1141. doi:10.1016/j.celrep.2013.10.014
52. Oldach P, Nieduszynski CA. Cohesin-mediated genome architecture does not define DNA replication timing domains. *Genes (Basel)*. 2019;10(3). doi:10.3390/genes10030196
53. Watson JD, Crick FHC. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*. 1953;171(4356):737-738. doi:10.1038/171737a0
54. Petryk N, Kahli M, D'Aubenton-Carafa Y, et al. Replication landscape of the human genome. *Nat Commun*. 2015;7:10208. doi:10.1038/ncomms10208
55. Fu H, Baris A, Aladjem MI. Replication timing and nuclear structure. *Curr Opin Cell Biol*. 2018;52:43-50. doi:10.1016/J.CEB.2018.01.004
56. Hadjadj D, Denecker T, Maric C, Fauchereau F, Baldacci G, Cadoret J-C. Characterization of the replication timing program of 6 human model cell lines. *Genomics Data*. 2016;9:113-117. doi:10.1016/J.GDATA.2016.07.003
57. Ryba T, Hiratani I, Sasaki T, et al. Replication Timing: A fingerprint for cell identity and pluripotency. Teichmann SA, ed. *PLoS Comput Biol*. 2011;7(10):e1002225. doi:10.1371/journal.pcbi.1002225
58. Pope BD, Ryba T, Dileep V, et al. Topologically associating domains are stable units of replication-timing regulation. *Nature*. 2014;515(7527):402-405. doi:10.1038/nature13986
59. Hansen RS, Thomas S, Sandstrom R, et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc Natl Acad Sci*. 2010;107(1):139-144. doi:10.1073/pnas.0912402107
60. Visser AE, Eils R, Jauch A, et al. Spatial distributions of early and late replicating chromatin in interphase chromosome territories. *Exp Cell Res*. 1998;243(2):398-407. doi:10.1006/EXCR.1998.4144
61. Dimitrova DS, Berezney R. The spatio-temporal organization of DNA replication sites is identical in primary, immortalized and transformed mammalian cells. *J Cell Sci*. 2002;115(21):4037-4051. doi:doi:10.1242/jcs.00087
62. Jackson DA, Pombo A. Replicon clusters are stable units of chromosome structure: Evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J Cell Biol*. 1998;140(6):1285-1295. doi:10.1083/JCB.140.6.1285
63. Nakayasu H, Berezney R. Mapping replicational sites in the eucaryotic cell nucleus. *J Cell Biol*. 1989;108(1):1-11. doi:10.1083/JCB.108.1.1
64. Xiang W, Julia Roberti M, Hériché JK, Huet S, Alexander S, Ellenberg J. Correlative live and super-resolution imaging reveals the dynamic structure of replication domains. *J Cell Biol*. 2018;217(6):1973-1984. doi:10.1083/jcb.201709074
65. Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289-293. doi:10.1126/science.1181369
66. Ryba T, Hiratani I, Lu J, et al. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res*. 2010;20(6):761-770. doi:10.1101/gr.099655.109
67. Moindrot B, Audit B, Klous P, et al. 3D chromatin conformation correlates

with replication timing and is conserved in resting cells. *Nucleic Acids Res*. 2012;40(19):9470-9481. doi:10.1093/nar/gks736
68. Yaffe E, Farkash-Amar S, Polten A, Yakhini Z, Tanay A, Simon I. Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. Bickmore WA, ed. *PLoS Genet*. 2010;6(7):e1001011. doi:10.1371/journal.pgen.1001011
69. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of chromosomal domains by loop extrusion. *Cell Rep*. 2016;15(9):2038-2049. doi:10.1016/j.celrep.2016.04.085
70. Nasmyth K. Disseminating the genome: Joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu Rev Genet*. 2001;35(1):673-745. doi:10.1146/annurev.genet.35.102401.091334
71. Wutz G, Várnai C, Nagasaka K, et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J*. 2017;36(24):3573-3599. doi:10.15252/embj.201798004
72. Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny LA. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci*. 2018;115(29):E6697-E6706. doi:10.1073/pnas.1717730115
73. Schwarzer W, Abdennur N, Goloborodko A, et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*. 2017;551(7678):51-56. doi:10.1038/nature24281
74. Lubelsky Y, Prinz JA, DeNapoli L, Li Y, Belsky JA, MacAlpine DM. DNA replication and transcription programs respond to the same chromatin cues. *Genome Res*. 2014;24(7):1102-1114. doi:10.1101/gr.160010.113
75. Nathanailidou P, Taraviras S, Lygerou Z. Chromatin and nuclear architecture: Shaping DNA replication in 3D. *Trends Genet*. July 2020. doi:10.1016/j.tig.2020.07.003
76. Dimitrova DS, Gilbert DM. The spatial position and replication timing of chromosomal domains are both established in early G1 phase. *Mol Cell*. 1999;4(6):983-993. doi:10.1016/S1097-2765(00)80227-0
77. Dileep V, Ay F, Sima J, Vera DL, Noble WS, Gilbert DM. Topologically-associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication timing program. *Genome Res*. 2015;25:1104-1113. doi:10.1101/gr.183699.114.4
78. Dixon JR, Jung I, Selvaraj S, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015;518(7539):331-336. doi:10.1038/nature14222
79. Lemaitre JM, Danis E, Pasero P, Vassetzky Y, Méchali M. Mitotic remodeling of the replicon and chromosome structure. *Cell*. 2005;123(5):787-801. doi:10.1016/j.cell.2005.08.045
80. Dileep V, Rivera-Mulia JC, Sima J, Gilbert DM. Large-scale chromatin structure-function relationships during the cell cycle and development: Insights from replication timing. *Cold Spring Harb Symp Quant Biol*. 2015;80:53-63. doi:10.1101/sqb.2015.80.027284
81. Mantiero D, Mackenzie A, Donaldson A, Zegerman P. Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *EMBO J*. 2011;30(23):4805-4814. doi:10.1038/emboj.2011.404
82. Natsume T, Müller CA, Katou Y, et al. Kinetochores coordinate

pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment. *Mol Cell*. 2013;50(5):661-674. doi:10.1016/j.molcel.2013.05.011
83. Fang D, Lengronne A, Shi D, et al. Dbf4 recruitment by forkhead transcription factors defines an upstream rate-limiting step in determining origin firing timing. *Genes Dev*. 2017;31(23-24):2405-2415. doi:10.1101/gad.306571.117
84. Sukackaite R, Cornacchia D, Jensen MR, et al. Mouse Rif1 is a regulatory subunit of protein phosphatase 1 (PP1). *Sci Rep*. 2017;7(1):1-10. doi:10.1038/s41598-017-01910-1
85. Hiraga S, Alvino GM, Chang FJ, et al. Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse Cdc7- mediated phosphorylation of the MCM complex. *Genes Dev*. 2014;28(4):372-383. doi:10.1101/gad.231258.113
86. Hiraga S, Ly T, Garzón J, et al. Human RIF 1 and protein phosphatase 1 stimulate DNA replication origin licensing but suppress origin activation . *EMBO Rep*. 2017;18(3):403-419. doi:10.15252/embr.201641983
87. Davé A, Cooley C, Garg M, Bianchi A. Protein Phosphatase 1 recruitment by Rif1 regulates DNA replication origin firing by counteracting DDK activity. *Cell Rep*. 2014;7(1):53-61. doi:10.1016/J.CELREP.2014.02.019
88. Mattarocci S, Shyian M, Lemmens L, et al. Rif1 controls DNA replication timing in yeast through the PP1 Phosphatase Glc7. *Cell Rep*. 2014;7(1):62-69. doi:10.1016/j.celrep.2014.03.010
89. Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov G V., Mirkin SM, Sunyaev SR. Human mutation rate associated with DNA replication timing. *Nat Genet*. 2009;41(4):393-395. doi:10.1038/ng.363
90. Woo YH, Li WH. DNA replication timing and selection shape the landscape of nucleotide variation in cancer genomes. *Nat Commun*. 2012;3(1):1-8. doi:10.1038/ncomms1982
91. De S, Michor F. DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. *Nat Biotechnol*. 2011;29(12):1103-1108. doi:10.1038/nbt.2030
92. Bester AC, Roniger M, Oren YS, et al. Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell*. 2011;145(3):435-446. doi:10.1016/j.cell.2011.03.044
93. Burrell RA, McClelland SE, Endesfelder D, et al. Replication stress links structural and numerical cancer chromosomal instability. *Nature*. 2013;494(7438):492-496. doi:10.1038/nature11935
94. Du Q, Bert SA, Armstrong NJ, et al. Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. *Nat Commun*. 2019;10(1):1-15. doi:10.1038/s41467-019-08302-1
95. Zhang J, Xu F, Hashimshony T, Keshet I, Cedar H. Establishment of transcriptional competence in early and late S phase. *Nature*. 2002;420(6912):198-202. doi:10.1038/nature01150
96. Lande-Diner L, Zhang J, Cedar H. Shifts in replication timing actively affect histone acetylation during nucleosome reassembly. *Mol Cell*. 2009;34(6):767-774. doi:10.1016/j.molcel.2009.05.027
97. Voichek Y, Bar-Ziv R, Barkai N. Expression homeostasis during DNA replication. *Science (80- )*. 2016;351(6277):1087-1090. doi:10.1126/science.aad1162

98. Müller CA, Nieduszynski CA. DNA replication timing influences gene expression level. *J Cell Biol*. 2017;216(7):1907-1914. doi:10.1083/jcb.201701061
99. Peters JM, Tedeschi A, Schmitz J. The cohesin complex and its roles in chromosome biology. *Genes Dev*. 2008;22(22):3089-3114. doi:10.1101/gad.1724308
100. Haering CH, Farcas AM, Arumugam P, Metson J, Nasmyth K. The cohesin ring concatenates sister DNA molecules. *Nature*. 2008;454(7202):297-301. doi:10.1038/nature07098
101. Sjögren C, Nasmyth K. Sister chromatid cohesion is required for postreplicative double-strand break repair in *Saccharomyces cerevisiae*. *Curr Biol*. 2001;11(12):991-995. doi:10.1016/S0960-9822(01)00271-8
102. Kim ST, Xu B, Kastan MB. Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage. *Genes Dev*. 2002;16(5):560-570. doi:10.1101/gad.970602
103. Kitagawa R, Bakkenist CJ, McKinnon PJ, Kastan MB. Phosphorylation of SMC1 is a critical downstream event in the ATM-NBS1-BRCA1 pathway. *Genes Dev*. 2004;18(12):1423-1438. doi:10.1101/gad.1200304
104. Rollins RA, Morcillo P, Dorsett D. Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics*. 1999;152(2):577-593. /pmc/articles/PMC1460629/?report=abstract. Accessed September 7, 2020.
105. Horsfield JA, Anagnostou SH, Hu JKH, et al. Cohesin-dependent regulation of Runx genes. *Development*. 2007;134(14):2639-2649. doi:10.1242/dev.002485
106. Kagey MH, Newman JJ, Bilodeau S, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 2010;467(7314):430-435. doi:10.1038/nature09380
107. Donze D, Adams CR, Rine J, Kamakaka RT. The boundaries of the silenced HMR domain in *Saccharomyces cerevisiae*. *Genes Dev*. 1999;13(6):698-708. doi:10.1101/gad.13.6.698
108. Wendt KS, Yoshida K, Itoh T, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*. 2008;451(7180):796-801. doi:10.1038/nature06634
109. Szabo Q, Bantignies F, Cavalli G. Principles of genome folding into topologically associating domains. *Sci Adv*. 2019;5(4):eaaw1668. doi:10.1126/sciadv.aaw1668
110. Vietri Rudan M, Barrington C, Henderson S, et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep*. 2015;10(8):1297-1309. doi:10.1016/j.celrep.2015.02.004
111. Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, Kanemaki M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods*. 2009;6(12):917-922. doi:10.1038/nmeth.1401
112. Natsume T, Kiyomitsu T, Saga Y, Kanemaki MT. Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. *Cell Rep*. 2016;15(1):210-218. doi:10.1016/J.CELREP.2016.03.001
113. Gilbert DM. Cell fate transitions and the replication timing decision point. *J Cell Biol*. 2010;191(5):899-903. doi:10.1083/jcb.201007125
114. Zuba-Surma EK, Kucia M, Abdel-Latif A, Lillard JW, Ratajczak MZ. *The ImageStream System: A Key Step to a New Era in Imaging*. Vol 45.; 2007.

doi:10.5603/4500
115. Sima J, Chakraborty A, Dileep V, et al. Identifying *cis* elements for spatiotemporal control of mammalian DNA replication. *Cell*. 2019;176(4):816-830.e18. doi:10.1016/j.cell.2018.11.036
116. Cremer M, Brandstetter K, Maiser A, et al. Cohesin depleted cells pass through mitosis and reconstitute a functional nuclear architecture. *bioRxiv*. October 2019:816611. doi:10.1101/816611
117. Boija A, Klein IA, Sabari BR, et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell*. 2018;175(7):1842-1855.e16. doi:10.1016/j.cell.2018.10.042
118. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A phase separation model for transcriptional control. *Cell*. 2017;169(1):13-23. doi:10.1016/j.cell.2017.02.007
119. Sansam CG, Pietrzak K, Majchrzycka B, et al. A mechanism for epigenetic control of DNA replication. *Genes Dev*. 2018;32(3-4):224-229. doi:10.1101/gad.306464.117
120. Yu J. Single-molecule studies in live cells. *Annu Rev Phys Chem*. 2016;67:565-585. doi:10.1146/annurev-physchem-040215-112451
121. Morrison A, Araki H, Clark AB, Hamatake RK, Sugino A. A third essential DNA polymerase in *S. cerevisiae*. *Cell*. 1990;62(6):1143-1151. doi:10.1016/0092-8674(90)90391-Q
122. Nick McElhinny SA, Gordenin DA, Stith CM, Burgers PMJ, Kunkel TA. Division of labor at the eukaryotic replication fork. *Mol Cell*. 2008;30(2):137-144. doi:10.1016/j.molcel.2008.02.022
123. Daigaku Y, Keszthelyi A, Müller CA, et al. A global profile of replicative polymerase usage. *Nat Struct Mol Biol*. 2015;22(3):192-198. doi:10.1038/nsmb.2962
124. Pursell ZF, Isoz I, Lundström EB, Johansson E, Kunkel TA. Yeast DNA polymerase ε participates in leading-strand DNA replication. *Science (80- )*. 2007;317(5834):127-130. doi:10.1126/science.1144067
125. Zhou ZX, Lujan SA, Burkholder AB, Garbacz MA, Kunkel TA. Roles for DNA polymerase δ in initiating and terminating leading strand DNA replication. *Nat Commun*. 2019;10(1):1-10. doi:10.1038/s41467-019-11995-z
126. Bulock CR, Xing X, Shcherbakova P V. DNA polymerase δ proofreads errors made by DNA polymerase ε. *Proc Natl Acad Sci*. 2020;117(11):6035-6041. doi:10.1073/pnas.1917624117
127. Zhou JC, Janska A, Goswami P, et al. CMG-Pol epsilon dynamics suggests a mechanism for the establishment of leading-strand synthesis in the eukaryotic replisome. *Proc Natl Acad Sci*. 2017;114(16):4141-4146. doi:10.1073/pnas.1700530114
128. Chilkova O, Jonsson BH, Johansson E. The quaternary structure of DNA polymerase ε from *Saccharomyces cerevisiae*. *J Biol Chem*. 2003;278(16):14082-14086. doi:10.1074/jbc.M211818200
129. Klinge S, Núñez-Ramírez R, Llorca O, Pellegrini L. 3D architecture of DNA Pol α reveals the functional core of multi-subunit replicative polymerases. *EMBO J*. 2009;28(13):1978-1987. doi:10.1038/emboj.2009.150
130. Ganai RA, Bylund G¨ OO, Johansson E. Switching between polymerase and exonuclease sites in DNA polymerase. *Nucleic Acids Res*. 2014;(1). doi:10.1093/nar/gku1353
131. Hogg M, Johansson E. DNA polymerase ε. *Subcell Biochem*. 2012;62:237-

257. doi:10.1007/978-94-007-4572-8_13
132. Araki H, Ropp PA, Johnson AL, Johnston LH, Morrison A, Sugino A. DNA polymerase II, the probable homolog of mammalian DNA polymerase epsilon, replicates chromosomal DNA in the yeast *Saccharomyces cerevisiae*. *EMBO J*. 1992;11(2):733-740. http://www.ncbi.nlm.nih.gov/pubmed/1537345. Accessed March 26, 2020.
133. Dua R, Levy DL, Campbell JL. Analysis of the essential functions of the C-terminal protein/protein interaction domain of *Saccharomyces cerevisiae* pol ε and its unexpected ability to support growth in the absence of the DNA polymerase domain. *J Biol Chem*. 1999;274(32):22283-22288. doi:10.1074/jbc.274.32.22283
134. Devbhandari S, Remus D. Rad53 limits CMG helicase uncoupling from DNA synthesis at replication forks. *Nat Struct Mol Biol*. April 2020:1-11. doi:10.1038/s41594-020-0407-7
135. Devbhandari S, Jiang J, Kumar C, Whitehouse I, Remus D. Chromatin constrains the initiation and elongation of DNA replication. *Mol Cell*. 2017;65:131-141. doi:10.1016/j.molcel.2016.10.035
136. Budd ME, Campbell JL. DNA polymerases δ and ε are required for chromosomal replication in Saccharomyces cerevisiae. *Mol Cell Biol*. 1993;13(1):496-505. http://mcb.asm.org/. Accessed June 8, 2020.
137. Navas TA, Zhou Z, Elledge SJ. DNA polymerase ε links the DNA replication machinery to the S phase checkpoint. *Cell*. 1995;80(1):29-39. doi:10.1016/0092-8674(95)90448-4
138. Kesti T, Flick K, Keränen S, Syväoja JE, Wittenberg C. DNA polymerase ε catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. *Mol Cell*. 1999;3(5):679-685. doi:10.1016/S1097-2765(00)80361-5
139. Lou H, Komata M, Katou Y, et al. Mrc1 and DNA polymerase ε function together in linking DNA replication and the S phase checkpoint. *Mol Cell*. 2008;32(1):106-117. doi:10.1016/j.molcel.2008.08.020
140. Ohya T, Kawasaki Y, Hiraga S, et al. The DNA polymerase domain of pol ε is required for rapid, efficient, and highly accurate chromosomal DNA replication, telomere length maintenance, and normal cell senescence in *Saccharomyces cerevisiae*. *J Biol Chem*. 2002;277(31):28099-28108. doi:10.1074/jbc.M111573200
141. Yeeles JTP, Janska A, Early A, Diffley JFX. How the eukaryotic replisome achieves rapid and efficient DNA replication. *Mol Cell*. 2017;65(1):105-116. doi:10.1016/j.molcel.2016.11.017
142. Morrison A, Bell JB, Kunkel TA, Sugino A. Eukaryotic DNA polymerase amino acid sequence required for 3' → 5' exonuclease activity. *Proc Natl Acad Sci*. 1991;88(21):9473-9477. doi:10.1073/pnas.88.21.9473
143. Shcherbakova P V., Pavlov YI, Chilkova O, Rogozin IB, Johansson E, Kunkel TA. Unique error signature of the four-subunit yeast DNA polymerase ε. *J Biol Chem*. 2003;278(44):43770-43780. doi:10.1074/jbc.M306893200
144. Meng X, Wei L, Devbhandari S, et al. DNA polymerase ε relies on a unique domain for efficient replisome assembly and strand synthesis. *Nat Commun*. 2020;11(1):1-13. doi:10.1038/s41467-020-16095-x
145. Winczura A, Appanah R, Tatham MH, Hay RT, De Piccoli G. The S phase checkpoint promotes the Smc5/6 complex dependent SUMOylation of Pol2, the catalytic subunit of DNA polymerase ε. *PLoS Genet*.

2019;15(11):e1008427. doi:10.1371/journal.pgen.1008427
146. Meng X, Wei L, Peng Id XP, Zhao X. Sumoylation of the DNA polymerase ε by the Smc5/6 complex contributes to DNA replication. 2019. doi:10.1371/journal.pgen.1008426
147. Grabarczyk DB, Silkenat S, Kisker C. Structural basis for the recruitment of Ctf18-RFC to the replisome. *Structure*. 2018;26(1):137-144.e3. doi:10.1016/j.str.2017.11.004
148. Edwards S, Li CM, Levy DL, Brown J, Snow PM, Campbell JL. *Saccharomyces cerevisiae* DNA polymerase ε and polymerase σ interact physically and functionally, suggesting a role for polymerase ε in sister chromatid cohesion. *Mol Cell Biol*. 2003;23(8):2733-2748. doi:10.1128/mcb.23.8.2733-2748.2003
149. Araki H, Hamatake RK, Johnston LH, Sugino A. DPB2, the gene encoding DNA polymerase II subunit B, is required for chromosome replication in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci*. 1991;88(11):4601-4605. doi:10.1073/pnas.88.11.4601
150. Araki H, Hamatake RK, Morrison A, Johnson AL, Johnston LH, Sugino A. Cloning DPB3, the gene encoding the third subunit of DNA polymerase II of *Saccharomyces cerevisiae*. *Nucleic Acids Res*. 1991;19(18):4867-4872. doi:10.1093/nar/19.18.4867
151. Ohya T, Maki S, Kawasaki Y, Sugino A. Structure and function of the fourth subunit (Dpb4p) of DNA polymerase ε in *Saccharomyces cerevisiae*. *Nucleic Acids Res*. 2000;28(20):3846-3852.
152. Cho RJ, Campbell MJ, Winzeler EA, et al. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell*. 1998;2(1):65-73. doi:10.1016/S1097-2765(00)80114-8
153. Pramila T, Wu W, Miles S, Noble WS, Breeden LL. The forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S-phase gap in the transcriptional circuitry of the cell cycle. *Genes Dev*. 2006;20(16):2266-2278. doi:10.1101/gad.1450606
154. Kudlicki A, Rowicka M, Otwinowski Z. SCEPTRANS: An online tool for analyzing periodic transcription in yeast. *Bioinformatics*. 2007;23(12):1559-1561. doi:10.1093/bioinformatics/btm126
155. Roseaulin LC, Noguchi C, Martinez E, Ziegler MA, Toda T, Noguchi E. Coordinated degradation of replisome components ensures genome stability upon replication stress in the absence of the replication fork protection complex. *PLoS Genet*. 2013;9(1):e1003213. doi:10.1371/journal.pgen.1003213
156. Roseaulin LC, Noguchi C, Noguchi E. Proteasome-dependent degradation of replisome components regulates faithful DNA replication. *Cell Cycle*. 2013;12(16):2564-2569. doi:10.4161/cc.25692
157. Daraba A, Gali VK, Halmai M, Haracska L, Unk I. Def1 promotes the degradation of Pol3 for polymerase exchange to occur during DNA-damage-induced mutagenesis in *Saccharomyces cerevisiae*. *PLoS Biol*. 2014;12(1). doi:10.1371/journal.pbio.1001771
158. Goswami P, Abid Ali F, Douglas ME, et al. Structure of DNA-CMG-Pol epsilon elucidates the roles of the non-catalytic polymerase modules in the eukaryotic replisome. *Nat Commun*. 2018;9(1):5061. doi:10.1038/s41467-018-07417-1
159. Langston LD, Zhang D, Yurieva O, et al. CMG helicase and DNA

polymerase ε form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proc Natl Acad Sci*. 2014;111(43):15390-15395. doi:10.1073/pnas.1418334111
160. Tercero JA, Diffley JFX. Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. *Nature*. 2001;412(6846):553-557. doi:10.1038/35087607
161. Lopes M, Foiani M, Sogo JM. Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Mol Cell*. 2006;21(1):15-27. doi:10.1016/j.molcel.2005.11.015
162. Georgescu RE, Langston L, Yao NY, et al. Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat Struct Mol Biol*. 2014;21(8):664-670. doi:10.1038/nsmb.2851
163. Sparks JL, Chistol G, Gao AO, et al. The CMG helicase bypasses DNA-protein cross-links to facilitate their repair. *Cell*. 2019;176(1-2):167-181.e21. doi:10.1016/j.cell.2018.10.053
164. Georgescu RE, Schauer GD, Yao NY, et al. Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation. *Elife*. 2015;2015(4). doi:10.7554/eLife.04988
165. Taylor MRG, Yeeles JTP. Dynamics of replication fork progression following helicase–polymerase uncoupling in eukaryotes. *J Mol Biol*. 2019;431(10):2040-2049. doi:10.1016/j.jmb.2019.03.011
166. Garbacz MA, Cox PB, Sharma S, Lujan SA, Chabes A, Kunkel TA. The absence of the catalytic domains of *Saccharomyces cerevisiae* DNA polymerase ε strongly reduces DNA replication fidelity. *Nucleic Acids Res*. 2019. doi:10.1093/nar/gkz048
167. Srivatsan A, Li B, Sanchez DN, et al. Essential *Saccharomyces cerevisiae* genome instability suppressing genes identify potential human tumor suppressors. *Proc Natl Acad Sci*. 2019;116(35):17377-17382. doi:10.1073/pnas.1906921116
168. Boddy MN, Russell P. DNA replication checkpoint. *Curr Biol*. 2001;11(23):R953-R956. doi:10.1016/s0960-9822(01)00572-3
169. Byun TS, Pacek M, Yee MC, Walter JC, Cimprich KA. Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes Dev*. 2005;19(9):1040-1052. doi:10.1101/gad.1301205
170. Zhou Z, Elledge SJ. DUN1 encodes a protein kinase that controls the DNA damage response in yeast. *Cell*. 1993;75(6):1119-1127. doi:10.1016/0092-8674(93)90321-G
171. Santocanale C, Diffley JFX. A Mec1-and Rad53-dependent checkpoint controls late-firing origins of DNA replication. *Nature*. 1998;395(6702):615-618. doi:10.1038/27001
172. Alcasabas AA, Osborn AJ, Bachant J, et al. Mrc1 transduces signals of DNA replication stress to activate Rad53. *Nat Cell Biol*. 2001;3(11):958-965. doi:10.1038/ncb1101-958
173. Gan H, Yu C, Devbhandari S, et al. Checkpoint kinase Rad53 couples leading- and lagging-strand DNA synthesis under replication stress. *Mol Cell*. 2017;68(2):446-455.e3. doi:10.1016/j.molcel.2017.09.018
174. Hogg M, Osterman P, Bylund GO, et al. Structural basis for processive DNA synthesis by yeast DNA polymerase epsilon. *Nat Struct Mol Biol*. 2014;21(1):49-55. doi:10.1038/nsmb.2712

175. Wade BO, Liu HW, Samora CP, Uhlmann F, Singleton MR. Structural studies of RFC$^{Ctf18}$ reveal a novel chromatin recruitment role for Dcc1. *EMBO Rep*. 2017;18(4):558-568. doi:10.15252/embr.201642825
176. García-Rodríguez LJ, De Piccoli G, Marchesi V, Jones RC, Edmondson RD, Labib K. A conserved Pol binding module in Ctf18-RFC is required for S-phase checkpoint activation downstream of Mec1. *Nucleic Acids Res*. 2015;43(18):8830-8838. doi:10.1093/nar/gkv799
177. Okimoto H, Tanaka S, Araki H, Ohashi E, Tsurimoto T. Conserved interaction of Ctf18-RFC with DNA polymerase ε is critical for maintenance of genome stability in *Saccharomyces cerevisiae*. *Genes to Cells*. 2016;21(5):482-491. doi:10.1111/gtc.12356
178. Shiomi Y, Masutani C, Hanaoka F, Kimura H, Tsurimoto T. A second proliferating cell nuclear antigen loader complex, Ctf18-replication factor C, stimulates DNA polymerase η activity. *J Biol Chem*. 2007;282(29):20906-20914. doi:10.1074/jbc.M610102200
179. Katou Y, Kanoh Y, Bando M, et al. S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature*. 2003;424(6952):1078-1083. doi:10.1038/nature01900
180. Lewis JS, Spenkelink LM, Schauer GD, et al. Single-molecule visualization of *Saccharomyces cerevisiae* leading-strand synthesis reveals dynamic interaction between MTC and the replisome. *Proc Natl Acad Sci*. 2017;114(40):10630-10635. doi:10.1073/pnas.1711291114
181. Szyjka SJ, Viggiani CJ, Aparicio OM. Mrc1 is required for normal progression of replication forks throughout chromatin in *S. cerevisiae*. *Mol Cell*. 2005;19(5):691-697. doi:10.1016/j.molcel.2005.06.037
182. Can G, Kauerhof AC, Macak D, Zegerman P. Helicase subunit Cdc45 targets the checkpoint kinase Rad53 to both replication initiation and elongation complexes after fork stalling. *Mol Cell*. 2019;73(3):562-573.e3. doi:10.1016/j.molcel.2018.11.025
183. Janke R, King GA, Kupiec M, Rine J. Pivotal roles of PCNA loading and unloading in heterochromatin function. *Proc Natl Acad Sci*. 2018;115(9):E2030-E2039. doi:10.1073/pnas.1721573115
184. Tackett AJ, Dilworth DJ, Davey MJ, et al. Proteomic and genomic characterization of chromatin complexes at a boundary. *J Cell Biol*. 2005;169(1):35-47. doi:10.1083/jcb.200502104
185. Yu C, Gan H, Serra-Cardona A, et al. A mechanism for preventing asymmetric histone segregation onto replicating DNA strands. *Science (80-)*. 2018;361(6409):1386-1389. doi:10.1126/science.aat8849
186. Iida T, Araki H. Noncompetitive counteractions of DNA polymerase ε and ISW2/yCHRAC for epigenetic inheritance of telomere position effect in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 2004;24(1):217-227. doi:10.1128/mcb.24.1.217-227.2004
187. Tsubota T, Tajima R, Ode K, et al. Double-stranded DNA binding, an unusual property of DNA polymerase ε, promotes epigenetic silencing in *Saccharomyces cerevisiae*. *J Biol Chem*. 2006;281(43):32898-32908. doi:10.1074/jbc.M606637200
188. Taylor MRG, Yeeles JTP. The initial response of a eukaryotic replisome to DNA damage. *Mol Cell*. 2018;70(6):1067-1080.e12. doi:10.1016/j.molcel.2018.04.022
189. Guilliam TA, Yeeles JTP. Reconstitution of translesion synthesis reveals a

mechanism of eukaryotic DNA replication restart. *Nat Struct Mol Biol*. April 2020:1-11. doi:10.1038/s41594-020-0418-4
190. Chilkova O, Stenlund P, Isoz I, et al. 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*. 2007;35(19):6588-6597. doi:10.1093/nar/gkm741
191. Lovett ST. Polymerase switching in DNA replication. *Mol Cell*. 2007;27:523-526. doi:10.1016/j.molcel.2007.08.003
192. Beattie TR, Kapadia N, Nicolas E, et al. Frequent exchange of the DNA polymerase during bacterial chromosome replication. *Elife*. 2017;6(e21763). doi:10.7554/eLife.21763
193. Lewis JS, Spenkelink LM, Jergic S, et al. Single-molecule visualization of fast polymerase turnover in the bacterial replisome. *Elife*. 2017;6(e23932). doi:10.7554/eLife.23932
194. Yuan Q, Dohrmann PR, Sutton MD, Mchenry CS. DNA polymerase III, but not polymerase IV, must be bound to a τ-containing DnaX complex to enable exchange into replication forks. *J Biol Chem*. 2016;291(22):11727-11735. doi:10.1074/jbc.M116.725358
195. Lewis JS, Spenkelink LM, Schauer GD, et al. Tunability of DNA polymerase stability during eukaryotic DNA replication. *Mol Cell*. 2020;77(1):17-25.e5. doi:10.1016/j.molcel.2019.10.005
196. Åberg C, Duderstadt KE, Van Oijen AM. Stability versus exchange: A paradox in DNA replication. *Nucleic Acids Res*. 2016;44(10):4846-4854. doi:10.1093/nar/gkw296
197. Geertsema HJ, van Oijen AM. A single-molecule view of DNA replication: the dynamic nature of multi-protein complexes revealed. *Curr Opin Struct Biol*. 2013;23(5):788-793. doi:10.1016/j.sbi.2013.06.018
198. Louvion JF, Havaux-Copf B, Picard D. Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. *Gene*. 1993;131(1):129-134. doi:10.1016/0378-1119(93)90681-R
199. McIsaac RS, Silverman SJ, McClean MN, et al. Fast-acting and nearly gratuitous induction of gene expression and protein depletion in *Saccharomyces cerevisiae*. Boone C, ed. *Mol Biol Cell*. 2011;22(22):4447-4459. doi:10.1091/mbc.e11-05-0466
200. Mans R, Van Rossum HM, Wijsman M, et al. CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res*. 2015;15:4. doi:10.1093/femsyr/fov004
201. Janke C, Magiera MM, Rathfelder N, et al. A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast*. 2004;21(11):947-962. doi:10.1002/yea.1142
202. Weinhandl K, Winkler M, Glieder A, Camattari A. Carbon source dependent promoters in yeasts. *Microb Cell Fact*. 2014;13(1):5. doi:10.1186/1475-2859-13-5
203. Mumberg D, Mulier R, Funk M. Regulatable promoters of *Saccharomyces cerevisiae*: Comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res*. 1994;22(25):5767-5768.
204. Raghuraman MK, Winzeler EA, Collingwood D, et al. Replication dynamics

of the yeast genome. *Science (80- )*. 2001;294(5540):115-121. doi:10.1126/science.294.5540.115
205. Yabuki N, Terashima H, Kitada K. Mapping of early firing origins on a replication profile of budding yeast. *Genes to Cells*. 2002;7(8):781-789. doi:10.1046/j.1365-2443.2002.00559.x
206. Alvino GM, Collingwood D, Murphy JM, Delrow J, Brewer BJ, Raghuraman MK. Replication in hydroxyurea: It's a matter of time. *Mol Cell Biol*. 2007;27(18):6396-6406. doi:10.1128/mcb.00719-07
207. Donaldson AD, Fangman WL, Brewer BJ. Cdc7 is required throughout the yeast S phase to activate replication origins. *Genes Dev*. 1998;12(4):491-501.
208. Newlon CS, Fangman WL. Mitochondrial DNA synthesis in cell cycle mutants of *Saccharomyces cerevisiae*. *Cell*. 1975;5(4):423-428. doi:10.1016/0092-8674(75)90061-6
209. Teytelman L, Thurtle DM, Rine J, Van Oudenaarden A. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci U S A*. 2013;110(46):18602-18607. doi:10.1073/pnas.1316064110
210. Lewis JS, Costa A. Caught in the act: Structural dynamics of replication origin activation and fork progression. *Biochem Soc Trans*. 2020;48(3):1057-1066. doi:10.1042/BST20190998
211. Kapadia N, El-Hajj ZW, Zheng H, Beattie TR, Yu A, Reyes-Lamothe R. Processive activity of replicative DNA polymerases in the replisome of live eukaryotic cells. *Mol Cell*. September 2020. doi:10.1016/j.molcel.2020.08.014
212. McVey M, Khodaverdian VY, Meyer D, Cerqueira PG, Heyer WD. Eukaryotic DNA polymerases in homologous recombination. *Annu Rev Genet*. 2016;50:393-421. doi:10.1146/annurev-genet-120215-035243
213. Yang W, Gao Y. Translesion and repair DNA polymerases: Diverse structure and mechanism. *Annu Rev Biochem*. 2018;87:239-261. doi:10.1146/annurev-biochem-062917-012405
214. Sima J, Chakraborty A, Dileep V, et al. Identifying *cis* elements for spatiotemporal control of mammalian DNA replication. *Cell*. 2018;176:1-15. doi:10.1016/j.cell.2018.11.036