SINGLE MOLECULE APPROACHES TO MAPPING DNA REPLICATION ORIGINS

A Master's Thesis Presented

By

VICTOR LIU

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

DECEMBER 26, 2017

SINGLE MOLECULE APPROACHES TO MAPPING DNA REPLICATION ORIGINS

A Master's Thesis Presented

By

VICTOR LIU

The signatures of the Master's Thesis Committee signify
completion and approval as to style and content of the Thesis

ANN RITTENHOUSE, Ph.D., Chair of Committee

SILVIA CORVERA, Ph.D., Member of Committee

WILLIAM ROYER, Ph.D., Member of Committee

NICK RHIND, Ph.D., Thesis Advisor

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that
the student has met all master's degree graduation requirements of the school.

ANTHONY CARRUTHERS, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

DECEMBER 26, 2017

DEDICATED TO THE GENERAL PUBLIC

for whom I serve

May the advancement of science and medicine

bring joy to your life.

Acknowledgements

The work presented in this thesis was made possible with the generous support and guidance from many people.

My Thesis Advisor, Dr. Nick Rhind, has played a key role in my education and my time in the laboratory. I thank him for the opportunity to learn and grow as a scientist. He has always maintained an open-door policy, and it has been a delight to see him actively engaged in many areas on campus. I admire Nick's bold approach in tackling some of the most challenging questions in the field of DNA replication, and his preference to innovate and to adopt new approaches to address some of the most fundamental and interesting questions in biology. I thank him for his generosity with his time, and for his scientific guidance. As he continues to advance the field of DNA replication, I look forward to learning more about his ideas and discoveries.

To graduate students in the Rhind Lab—Dan Keifenheim, Melissa Greven, Livio Dukaj, Makoto Ohira, and Divya Iyer—thank you for your support and advice. Even though our projects had little overlap, being in a small lab with only 4-5 graduate students made it easy to know each other. I extend my appreciation to Divya, who taught me numerous methods and allowed me to bounce ideas about my projects. To Shankar Das, former postdoctoral fellow, and Ryan Holmes, former undergraduate student, thank you for helping with my projects.

Faculty members have been instrumental in my professional development. I extend my deepest thanks to Dr. Gyongyi Szabo and Dr. Silvia Corvera for their valuable advice. I thank Dr. Sharon Cantor, Dr. Craig Peterson, Dr. Michael Volkert, Dr. Michael Brodsky, and Dr. Dannel McCollum for their scientific insight, advice, and feedback. I thank Dr. Ann Rittenhouse, Dr. Silvia Corvera, and Dr. William Royer for providing feedback, and Anne Michelson for her kindness and assistance.

The Biochemistry and Molecular Pharmacology (BMP) Department and the University of Massachusetts Medical School is a wonderful place to learn. To Judy Mwangi, Josh Gardner, Luca Leone, and other staff members in the department, thank you for your help. I thank Ornella for her smile and for making media. To Ivery and Victor, thank you for keeping our laboratory meticulously clean. I extend my deepest appreciation to Irene, who has helped our laboratory in maintaining a fresh supply of clean glassware, especially after a busy weekend of experiments. To the person who prefers to remain anonymous, thank you for sharing your delicious home-cooked meals, which I shall dearly miss.

To my family, thank you for your unconditional support and for teaching me the joy of helping other people. To classmates and friends, I am particular grateful for the long lasting friendships that many of us have formed. I am amazed by how fast life goes by. Yet, I am amazed at how relationships can withhold the test of time.

Finally, I wish to thank myself, for keeping life in perspective and for being persistent. Without you, none of this would have been possible.

Abstract

DNA replication is a fundamental process that is primarily regulated at the initiation step. In higher eukaryotes, the location and properties of replication origins are not well understood. Existing genome-wide approaches to map origins—such as nascent strand abundance mapping, Okazaki fragment mapping, or chromatin immunoprecipitation-based assays—average the behavior of a population of cells. However, due to cell-to-cell variability in origin usage, single molecule techniques are necessary to investigate the actual behavior of a cell. Here, I investigate the feasibility of using three single molecule, genome-wide technologies to map origins of replication. The Pacific Biosciences Single Molecule Real-Time (SMRT) sequencing technology, the BioNano Genomics Irys optical mapping technology, and the Oxford Nanopore Technologies MinION nanopore sequencing technology are promising approaches that can advance our understanding of DNA replication in higher eukaryotes.

Table of Contents

List of Tables

List of Figures

List of Third Party Copyrighted Material

Figure Number	Publisher	Licence Number
Figure 2.1.	The American Association for the Advancement of Science	4116581317407
Figure 2.4.	Nature Publishing Group	4116570441742
Figure 4.1.	Elsevier	4159481350131

The following figures were reproduced from Pacific Biosciences with permission:

Figure 2.2.	Pacific Biosciences
Figure 2.3.	Pacific Biosciences

The following figure was reproduced from a journal and no permission was required:

Figure 4.3.	National Academy of Sciences

List of Abbreviations

AmdU	5-azidomethyl-2'-deoxyuridine
AmdUTP	5-azidomethyl-2'-deoxyuridine triphosphate
ACS	Autonomous replicating sequence (ARS) consensus sequence
ARS	Autonomous replicating sequence
BCN	Bicyclo[6.1.0]nonyne
BCN-OH	Bicyclo[6.1.0]non-4-yn-9-ylmethanol
BrdU	5-bromo-2'-deoxyuridine
BrdUTP	5-bromo-2'-deoxyuridine triphosphate
BTTP	3-(4-((bis((1-*tert*-butyl-1*H*-1,2,3-triazol-4-yl)methyl)amino)methyl)-1*H*-1,2,3-triazol-1-yl)propanol
ChIP	Chromatin immunoprecipitation
CuAAC	Copper(I)-catalyzed azide-alkyne cycloaddition
DBCO	Dibenzocyclooctyne or azadibenzocyclooctyne
DBCO-OH	Dibenzocyclooctyne acid
DBCO-PEG4-TAMRA	Dibenzylcyclooctyne-PEG4-5/6-Tetramethylrhodamine
DIBO	Dibenzocyclooctyne
DIFO	Difluorocyclooctyne
DMEM	Dubecco's Minimum Essential Media
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPBS	Dubecco's Phosphate Buffered Saline
dsDNA	Double-stranded DNA

dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dNTP	Deoxynucleotide triphosphate
dTTP	Deoxythymidine triphosphate or thymidine triphosphate
dUTP	Deoxyuridine triphosphate
EDTA	Ethylene Diamine Triacetic Acid
EdU	5-ethynyl-2'-deoxyuridine
EdUTP	5-ethynyl-2'-deoxyuridine triphosphate
FISH	Fluorescent in situ hybridization
GINS	GO-ICHI-NI-SAN
HU	Hydroxyurea
IdU	5-iodo-2'-deoxyuridine
IdUTP	5-iodo-2'-deoxyuridine triphosphate
IPD	Interpulse duration or interpulse distance
LB	Lysogeny broth
MCM	Minichromosome maintenance
MEL-HSV-tk	Mouse erythroleukemia cells with herpes simplex virus thymidylate synthase gene integration.
MMS	Methyl methanesulfonate
NFR	Nucleosome free region
ORC	Origin recognition complex
PacBio	Pacific Biosciences
PCR	Polymerase chain reaction
Pre-RC	Pre-recognition complex
RIP mapping	Replication initiation point mapping
RPA	Replication protein A

SMRT sequencing	Single molecule real time sequencing
SPAAC	Strain-promoted azide-alkyne cycloaddition
ssDNA	Single-stranded DNA
T	Thymidine
TE	Tris-EDTA
USB	Universal Serial Bus
YOYO-1	{1,1'-(4,4,8,8-tetramethyl-4,8-diazaundecamethylene)bis[4-[(3-methylbenzo-1,3-oxazol-2-yl)methylidene]-l,4-dihydroquinolinium] tetraiodide}
ZMW	Zero-mode waveguide

Preface

The work presented in this thesis was made possible with the generous help and expertise of other scientists. Contributors are noted below. The author extends his deepest appreciation for many other people who have worked behind the scenes to make the studies possible.

In addition to the projects presented in this thesis, the author also pursued other DNA replication studies. The investigation of novel nucleoside analogs for labeling replication tracks *in vivo*, and the two dimensional gel electrophoresis studies to assess origin activity, are among the work not included in this thesis but have nonetheless been part of the author's training.

Chapter II – List of Contributors

V. Liu conceived the project in Chapter II, designed the study, performed the experiments, and analyzed the data. N. Rhind provided guidance and feedback. V. Liu would like to thank Tyler Borrman (UMMS) and Richard Hall (Pacific Biosciences), Michael Weiand (Pacific Biosciences), and John Harting (Pacific Biosciences) for guidance in extracting and analyzing the data.

Chapter III – List of Contributors

The work in Chapter III was made possible with the generous assistance of members at BioNano Genomics (San Diego, CA), including Alex Hastie, Jeff Reifenberger, Denghong Zhang, and Han Cao.

N. Rhind conceived the overall project and suggested the EdU labeling approach and the EdU with BTTP ligand approach. V. Liu conceived the DNA fluorescence approach and the labeling of the replication forks. V. Liu, N. Rhind, H. Cao, A. Hastie, J. Reifenberger, D.H. Zhang, participated in the design of experiments. Ryan Holmes synthesized 5-azidomethyl-2'-deoxyuridine (AmdU) at UMMS prior to the compound being commercially available. V. Liu performed all experiments, except for DNA processing and mapping, which were done by our collaborators at BioNano Genomics. V. Liu, Tyler Borrman (UMMS), J. Reifenberger, A. Hastie, D.H. Zhang, and N. Rhind participated in the data analysis. V. Liu prepared all figures in Chapter III, except for Figures 3.6 and 3.7, which were prepared by N. Rhind and modified for this publication.

V. Liu would like to thank Professor Tom Brown (Oxford University) and members of his laboratory for supplying 5-azidomethyl-2'-deoxyuridine triphosphate (AmdUTP), and Dr. Peng Wu for supplying the ligand BTTP: (3-(4-((bis((1-*tert*-butyl-1*H*-1,2,3-triazol-4-yl)methyl)amino)methyl)-1*H*-1,2,3-triazol-1-yl)propanol).

Chapter IV – List of Contributors

The project in Chapter IV is a collaboration to map sites of DNA replication via nanopore sequencing. The project was conceived and is led by John Urban in Dr. Susan Gerbi's Laboratory (Brown University). V. Liu, J. Urban, S. Gerbi, and N. Rhind participated in the design of the wet experiments. V. Liu performed and optimized the experiments to obtain barcoded DNA with modified bases. J. Urban processed the DNA samples for nanopore sequencing. J. Urban and other collaborators analyzed the data. The reader is advised to refer to J. Urban's dissertation for details on each member's contributions:

> Urban, John, "The genome and DNA puff sequences of the fungus fly, Sciara coprophila, and genome-wide methods for studying DNA replication" (2017). *Molecular Biology, Cell Biology, and Biochemistry Theses and Dissertations.* Brown Digital Repository. Brown University Library. https://repository.library.brown.edu/studio/item/bdr:733543/

Chapter V – List of Contributors

The project in Chapter V stems from previous work in the laboratory (Willis and Rhind, 2009a, 2010, 2011; Willis et al., 2016). N. Rhind conceived the project and provided guidance and feedback. V. Liu performed the experiments, collected the data, analyzed the data, and prepared the figures. V. Liu would like to thank N. Willis and D. Iyer for technical assistance and advice on the project.

Chapter I: Introduction to Origins of Replication

1.1. Introduction

The oldest known fossil of our species, *Homo sapiens*, was recently discovered at the Jebel Irhoud site in Morocco (Hublin et al., 2017). Together with Middle Stone Age artefacts discovered at the site, there is now evidence that our species dates back to 300,000 years ago (Richter et al., 2017). Although our species has perpetuated through many generations, it is only in recent times that we have begun to investigate the biological mechanisms that drive the survival of our species. The human body is made of organs, tissues, and cells, which are fundamentally formed from building blocks encoded by deoxyribonucleic acid (DNA). A key biological process necessary for the continual existence of our species is thus DNA replication.

DNA replication is the process by which cells synthesize DNA and duplicate the genetic sequence. In eukaryotes, DNA replication occurs during the synthesis phase (S-phase) of the cell cycle. Since DNA replication is the means by which genetic information is transmitted to daughter cells, any misregulation of DNA replication in metazoans can lead to detrimental consequences such as developmental abnormalities and oncogenesis.

1.1.1. Semiconservative model of DNA replication

The discovery of deoxyribonucleic acid (DNA) as hereditary material carrying genetic information and its structure led to investigations on DNA replication (Avery et al., 1944; Hershey and Chase, 1952; Watson and Crick, 1953a). One month after their landmark report on the double helical structure of DNA, Watson and Crick proposed the semiconservative model of DNA replication (Watson and Crick, 1953b). In this model, replication occurs with the melting of two parental DNA strands that serve as templates for base-pairing of nucleotides. The nucleotides are then polymerized to form the daughter strand.

In an elegant experiment using isotopic labeling of nitrogen in purines and pyrimidines, Meselson and Stahl tested whether DNA replication occurs through a conservative, semiconservative, or dispersive mechanism (Meselson and Stahl, 1958). In this experiment, *Escherichia coli* cells were grown for fourteen generations in medium containing ^{15}N (heavy nitrogen isotope). This approach yielded parental DNA consisting of two heavy strands. Exponentially growing cells were then shifted to medium containing ^{14}N (light nitrogen isotope) and allowed to replicate. Using CsCl density gradient centrifugation, the authors showed that after 1 replication cycle, only one DNA band was present and its density was intermediate between the control heavy DNA and the control light DNA. After the second replication cycle, there were two DNA bands: one with the same intermediate density from the previous cycle, and one with density of the control light DNA. Their results indicated that DNA replication occurs in a semiconservative manner.

Semiconservative replication was first reported on a chromosomal level in the eukaryotic bean *Vicia faba* one year prior to Meselson and Stahl's report (Taylor et al., 1957). In this study, bean roots were grown in nutrients containing ^{3}H-thymidine to label DNA, and labeled metaphase chromosomes were visualized on autoradiographs. After growing to the point where the metaphase chromosomes showed incorporation of ^{3}H-thymidine, nutrients were then changed to contain non-radioactive thymidine. Prior to the nutrient change, labels were present in both chromatids of all metaphase chromosomes. After the first replication cycle, both chromatids of all metaphase chromosomes had half the original label content. After the second replication cycle, there were two populations of chromosomes: one population had half the original label content (from the previous cycle), and one population was unlabeled. These findings thus provided evidence of semiconservative replication, albeit on a chromosomal level.

The discovery of semiconservative DNA replication led scientists to investigate the mechanisms that govern DNA replication. The bacterium *E. coli* became a model organism for many replication studies. In *E. coli*, it was found that initiation of new rounds of DNA replication, but not the completion of ongoing rounds of replication, required *de novo* protein synthesis (Maaløe and Hanawalt, 1961). This study involved the assessment of ^{14}C-thymidine uptake to determine DNA content in the context of amino acid deprivation. The same conclusion was reached with a slightly more sophisticated approach using a density label (5-bromouracil, which has a higher molecular weight than uracil) and a radioactive label (^{3}H-thymidine) in the context of amino acid starvation (Lark *et al.* 1963). Moreover, *E. coli* has one origin of replication—that is, one point at

which replication begins (Bonhoeffer and Gierer, 1963). DNA fiber autoradiographic studies of ^{3}H-thymidine labeled DNA showed that the *E. coli* chromosome is circular (Cairns, 1963). Genetic marker frequency analyses in synchronized *E. coli* and in transformed *Bacillus subtilis* provided support that bacterial DNA replication—once initiated—proceeds sequentially with directionality along the DNA template (Nagata, 1963; Yoshikawa and Sueoka, 1963). Many of these studies led to the proposal of the replicon model to explain DNA replication in *E. coli* (Jacob et al., 1963).

1.1.2. Replicon model of DNA replication

The early observations on bacterial replication provided the framework for the replicon model of DNA replication in *E. coli* (Jacob et al., 1963). In this model, DNA replication involves a *trans*-acting factor ("initiator") that binds to a *cis*-acting DNA sequence ("replicator"). This event is necessary to drive replication initiation at the replication start site ("origin", as coined later by (Huberman and Riggs, 1968)). The model posits that DNA replication occurs in units ("replicons"), with each unit being the region that is replicated from an initiator-replicator pair. The initiator was proposed to act on the replicator as a positive regulatory factor, unlike the negative inducible system of the *lac operon* where a repressor binds to inhibit gene transcription in the absence of lactose (Jacob and Monod, 1961). The replicator sequence can overlap with the origin (Bell, 2002), and the terms are sometimes used interchangeably (Stillman, 1993), but the two terms differ (Bell, 2016): the replicator is the *cis*-acting sequence that the initiator acts on and is the sequence required for replication activity (Jacob et al., 1963), whereas the origin is the specific site at which initiation begins (Huberman and Riggs, 1968). Implicit in the replicon model is that DNA sequence is a major determinant of origin activation. This notion led to the search for replicator sequences and origins, which still continues today.

1.1.3. Origin of replication in *Escherichia coli*

The replicon model of DNA replication is now widely accepted to hold true for *E. coli*. The *E. coli* K-12 genome consists of one circular 4.6Mbp chromosome with only one replicator. DNA replication initiates from a single site and proceeds bidirectionally (Bird et al., 1972; Masters and Broda, 1971; Prescott and Kuempel, 1972; Yahara, 1972). Since the *E. coli* genome has only one origin of replication, there is one replicon and it encompasses the entire chromosome. Replication initiation is dependent upon the activity of DnaA, the initiator protein, which acts on *oriC*, the replicator sequence (Bramhill and Kornberg, 1988).

The origin of replication, *oriC*, is located at the 3.92Mb position (Blattner et al., 1997; Marsh, 1978; Marsh and Worcel, 1977; Yasuda and Hirota, 1977). The origin is a 232bp-245bp sequence that contains multiple 9-mer repeats (DnaA boxes) and three 13-mer AT-rich repeats (DNA unwinding element; DUE) (Chakraborty et al., 1982; Tabata et al., 1983).

DnaA, the initiator, is an AAA+ type of ATPase that can bind ATP or ADP (Sekimizu et al., 1987). Both DnaA-ATP and Dna-ADP can bind the origin of replication at the DnaA boxes (5'-TTATNCACA-3', where N = any canonical nucleotide) (Fuller et al., 1984; Matsui et al., 1985; Sekimizu et al., 1987). However, it is the binding of the active form, DnaA-ATP, that ultimately leads to DNA unwinding at the AT-rich DUE repeats (Sekimizu et al., 1987). As DnaA binds the origin, it undergoes self-oligomerization and forms a nucleoprotein filament. The DnaA oligomers, combined with negatively supercoiled DNA, induce a localized unwinding of duplex DNA. This

process generates single-stranded DNA for helicase loading. DnaA recruits DnaB, the replicative helicase, in the form of the DnaB-DnaC complex. DnaC, the helicase loader, facilitates this process via the DnaC-ssDNA interaction. This leads to replisome assembly and bidirectional replication (Kaguni, 2006; Mott and Berger, 2007; Skarstad and Katayama, 2013; Zakrzewska-Czerwinska et al., 2007).

1.1.4. Origins of replication in *Saccharomyces cerevisiae*

The replicon model explained chromosomal DNA replication in *E. coli* and other prokaryotes. In this section, I shall discuss whether the classic replicon model is applicable to eukaryotes. Unlike prokaryotes, eukaryote genomes have multiple chromosomes. Moreover, DNA fiber autoradiography showed the presence of multiple sites of DNA replication initiation along a eukaryotic chromosome (Huberman and Riggs, 1968). If the replicon model held true in eukaryotes, then what are the initiators and replicators that govern DNA replication?

Early studies in *S. cerevisiae*, a unicellular eukaryote, identified putative replicator sequences by their ability to maintain plasmid stability (Beach et al., 1980; Chan and Tye, 1980; Stinchcomb et al., 1979). In this assay, fragments of genomic DNA were cloned into plasmids with selectable markers, transformed into yeast, and screened for their ability to support plasmid DNA maintenance and replication. The identified sequences were termed autonomously replicating sequences (ARS) and had common sequence features. The budding yeast ARS is ~100bp and contains an ARS consensus sequence (ACS) that is necessary but not sufficient for replication initiation.

The canonical ACS is an 11-bp AT-rich motif with the sequence 5'-(A/T)TTTA(T/C)(A/G)TTT(A/T)-3', but the consensus sequence can be expanded to a 17-bp motif at other replicators like ARS309 (Theis and Newlon, 1997). The B1, B2, and B3 elements also contribute to ARS activity. The ACS and B1 elements serve as binding sites for the initiator, the origin recognition complex (ORC) (Bell and Stillman, 1992; Rao et al., 2012; Rowley et al., 1995). The B2 element facilitates in the loading of the MCM helicases after ORC binding (Lipford and Bell, 2001; Zou and Stillman, 2000). The B3 element is the binding site for ARS-binding factor 1 (Abf1), a transcription factor that facilitates nucleosome exclusion at the ACS and enhances replication (Diffley and Stillman, 1988; Marahrens and Stillman, 1992; Venditti et al., 1994).

The ARS plasmid assay yielded a list of putative origins in a non-chromosomal context. The development of two-dimensional (2D) gel electrophoresis of replication intermediates allowed the verification of origin activity (Brewer and Fangman, 1987; Huberman et al., 1987). At an origin, DNA replication initiation leads to the unwinding of duplex DNA and the synthesis of nascent daughter-strand DNA. This process forms a replication bubble structure at the origin, which can be detected via 2D gel electrophoresis. In this approach, plasmid DNA or chromosomal DNA are digested with a restriction enzyme that targets sites flanking the putative origin. The digested DNA is loaded onto an agarose gel and separated in the first dimension based on size. Next, the gel lane is excised, rotated 90^{o}, embedded into a new agarose gel, and electrophoresed. DNA is separated in the second dimension based on size and shape. DNA is then transferred to a membrane and a Southern blot is performed using a radioactive probe

specific for the locus of interest. Replication intermediates such as replication bubbles and replication forks have different shapes and thus can be resolved on 2D gels. The detection of a replication bubble is confirmation of origin activity. Many, but not all, of the ARS identified in the plasmid assay were shown to have chromosomal origin activity using 2D gel electrophoresis (Deshpande and Newlon, 1992; Dubey et al., 1991).

The consensus sequences in *S. cerevisiae* replicators facilitated the identification of the origin recognition complex (ORC), a heterohexameric protein with ATPase activity that functions as the initiator (Bell and Stillman, 1992). DNA replication initiation in *S. cerevisiae* is a two-step process, and the reader is advised to consult detailed reviews for primary references for the remainder of this paragraph (Bell, 2002; Bell and Dutta, 2002; Fragkos et al., 2015; Kelly and Stillman, 2006). The first step—origin licensing—occurs in G1 phase when ORC binds DNA and recruits the accessory proteins Cdc6 and Cdt1 to load the replicative helicase, MCM2-7. The MCM complex is loaded as a double hexamer on double-stranded DNA and remains inactive in G1 phase. Together, these essential proteins form an assembly called the pre-replication complex (pre-RC). The pre-RCs—with the inactive MCM complexes—act as potential origins that can be activated in the subsequent S phase. The second step—origin activation—occurs during S phase, when Cdc45 and GINS are recruited to the loaded inactive MCM complexes to form the active Cdc45/MCM2-7/GINS (CMG) helicase. The recruitment of Cdc45 and GINS is driven by cyclin-dependent kinases (CDKs) and Dbf4-dependent kinases (DDKs). The active CMG complex unwinds the double-stranded DNA to allow DNA polymerases to start bidirectional replication.

The temporal separation of origin licensing (G1 phase) and origin activation (S phase) ensures that replication does not occur more than once per cell cycle. Re-replication in *S. cerevisiae* may prevented through multiple mechanisms, including steric hindrance of ORC, CDK-mediated phosphorylation of ORC, downregulation of Cdc6, and nuclear exclusion of Cd1t and MCM2-7 (Chen and Bell, 2011; Dahmann et al., 1995; Drury et al., 2000; Liang and Stillman, 1997; Nguyen et al., 2001; Vas et al., 2001; Wilmes, 2004).

1.1.5. Nucleosome positioning and origin activity

Although replicators in *S. cerevisiae* were found to have consensus DNA sequences, the same principle does not appear to apply in higher eukaryotes. The lack of common replicator DNA sequences in higher eukaryotes suggests that higher order structures—such as chromatin organization—may play a role in defining origins. In eukaryotes, there is emerging evidence that a major determinant of origin activity is nucleosome positioning. Using a TRP1ARS1 plasmid in *S. cerevisiae*, it was found that nucleosomes positioned stably in the ARS1 locus (Thoma et al., 1984). Another study showed that the ACS of the ARS1 origin contains a nucleosome-free region (NFR) (Simpson, 1990). The nucleosome-free region at this ACS is important for origin activity, because a nucleosome placed into the ACS region—via plasmid mutations—decreased plasmid stability (Simpson, 1990). This finding may be due to the impediment of ORC to access its DNA binding site. ORC is required for positioning nucleosomes proximal to the ARS1 origin (Lipford and Bell, 2001). Nucleosome positioning adjacent to the ARS1 origin is important for its pre-replicative complex formation (Lipford and Bell, 2001).

When the proximal nucleosome is positioned further away—such that the NFR increases—there is decreased plasmid stability (Lipford and Bell, 2001). Moreover, MCM2-7 loading on the ARS1 origin is dependent on nucleosome position (Lipford and Bell, 2001). These findings suggest that the nucleosome positioning flanking the ACS, and the interaction between ORC and the nucleosome(s), are important determinants of origin activity.

In *S. cerevisiae*, the number of ACS predicted sequences—potential ORC binding sites—is greater than the number of origins by as much as two orders of magnitude (Belsky et al., 2015; Eaton et al., 2010). Using ChIP-seq, Eaton and colleagues identified ORC-bound ACS and compared their flanking nucleosome positioning and occupancy to non-ORC bound ACS. Their results showed that ORC-bound ACS are located in nucleosome-free regions. Moreover, the nucleosomes flanking the ORC-bound ACS are well-positioned and their distance from the ACS is conserved. This pattern is consistent with an analysis of nucleosome positioning (Berbenetz et al., 2010) in a list of *S. cerevisiae* origins (Nieduszynski, 2006) combined with nucleosome positioning data from a microarray study (Lee et al., 2007). The importance of nucleosome positioning on origin activity was further suggested in a genome-wide mapping of nucleosomes. Using micrococcal nuclease (MNase) to digest linker DNA between nucleosomes, followed by sequencing of residual DNA, a correlation was found between nucleosome position and MCM2-7 complex ChIP-seq signal (Belsky et al., 2015). This raises the possibility that loaded MCM2-7 helicase may interact with nucleosomes.

Nucleosome positioning was found to correlate with origin activation time. Origins that fired early generally had wider NFRs than origins that fired late (Soriano et al., 2014). In a recent study, nucleosome occupancy, as determined from histone H3-ChIP seq, was found to be reduced at ORC-bound ACS compared to ORC-unbound ACS (Rodriguez et al., 2017). Additionally, lower nucleosome occupancy correlated with earlier origin activation times and a higher probability of origin firing (Rodriguez et al., 2017). Together, these studies underscore the interplay between nucleosome occupancy and ORC, and suggest that both are important for determining origin location and origin properties.

The importance of nucleosome positioning to origin activity is also supported by the finding that ARS-Binding Factor 1 (ABF1) binds to the B3 element of ARS (Diffley and Stillman, 1988; Marahrens and Stillman, 1992) and acts as a physical boundary to facilitate nucleosome exclusion at the ACS (Venditti et al., 1994). Reconstituted DNA replication initiation assays *in vitro* have provided evidence that local nucleosome positioning is important for origin activation and replication fork progression (Azmi et al., 2017; Devbhandari et al., 2017; Kurat et al., 2017).

Besides *S. cerevisiae* (Mavrich et al., 2008), the tendency for origins to be in a nucleosome-free region is also evident in *Drosophila* (MacAlpine et al., 2010), and mammalian cells (Cayrou et al., 2015; Lubelsky et al., 2011; Yin et al., 2009). In *S. pombe*, there is evidence to suggest that origins tend to be in regions with low nucleosome occupancy regions (Givens et al., 2012; Lantermann et al., 2010; Xu et al., 2012), although one study did not find a correlation (De Castro et al., 2012).

1.1.6. Origins of replication in *Schizosaccharomyces pombe*

Given that *S. cerevisiae* origins were identified with replicators that had specific sequences, the replicon model continued to form the framework for investigating replication in other eukaryotes. The ARS plasmid assay was performed to identify replicator sequences in another unicellular eukaryote, the fission yeast *Schizosaccharomyces pombe*. However, unlike *S. cerevisiae*, the ARS elements identified in *S. pombe* lack a consensus nucleotide sequence (Wright et al., 1986). The *S. pombe* ARS elements tend to be long (>500bp) and highly AT-rich (Dai et al., 2005). The *S. pombe* ARS activity was retained when the elements were replaced with unrelated AT-rich sequences (Okuno et al., 1999). Based on this, it was hypothesized that the AT-rich regions serve as replicator sequences. Consistent with this hypothesis, a subunit of the *S. pombe* ORC initiator, Orc4p, was found to have AT-hook motifs that interact with AT-rich sequences (Chuang and Kelly, 1999).

1.1.7. Origins of replication in higher eukaryotes

In metazoans, where the genomes are more complex than in yeast, the search for replicators that drive DNA initiation produced results that were less clear. DNA injected into *Xenopus* egg extracts (*in vitro*) initiated replication and completed semiconservative replication, regardless of the DNA sequence (Harland and Laskey, 1980; Méchali and Kearsey, 1984). The ARS plasmid assay has not been helpful in identifying replicator sequences in mammalian cells. Plasmids with human DNA inserts were maintained independent of the insert sequence (Krysan and Calos, 1991). Studies suggest that the size of the human genomic fragments, and not their sequence, affects plasmid

maintenance (Caddie and Calos, 1992; Heinzel et al., 1991). Moreover, the complexity of the nuclear organization in higher eukaryotes raises the concern that any identified replicator sequences in plasmids may not fully recapitulate the behavior of chromosomal origins.

The complexity of the replication system in higher eukaryotes is exemplified by the replicons at the *Drosophila melanogaster* chorion gene locus. The chorion genes are amplified in the ovarian follicle cells to facilitate the rapid production of eggshell proteins (Orr-Weaver, 1991). The amplification process involves multiple rounds of replication initiation, which increases the gene copy number. Structurally, the repeated rounds of replication initiation form concentric layers of replication bubbles that resemble an "onionskin" (Claycomb et al., 2002; Osheim et al., 1988). DNA replication initiation can occur at multiple sites in the region, but in 70-80% of the time, it occurs at a site called ori-beta (Heck and Spradling, 1990). Replication at this locus requires the 320-bp amplification control element 3 (ACE3), which is located 1.5kb upstream of ori-beta (Carminati et al., 1992; Delidakis and Kafatos, 1989; Heck and Spradling, 1990). The *Drosophila* initiator, ORC, binds ACE3 *in vivo* (Austin et al., 1999). When an insulator was placed between the ACE3 and ori-beta elements, amplification was inhibited (Lu et al., 2001). Together, these findings suggest that ACE3 serves as a replicator that facilitates origin activation at the locus, primarily at ori-beta. This locus thus represents a scenario in which the replicator and the origin are distinct (Stillman, 1993).

Despite the low utility of using the ARS plasmid assay to identify metazoan replicators, a handful of mammalian origins have been identified using other techniques.

Among the well-studied metazoan origins is the replication start site at the human lamin B2 locus. The genomic segment containing the origin at the human lamin B2 locus was first isolated using a density labeling approach (Tribioli et al., 1987). In this study, synchronized human HL60 cells were released into S-phase in the presence of tritiated 5-bromo-2'-deoxyuridine (^{3}H-BrdU), which is an analog of thymidine with a higher molecular weight. The newly replicated DNA containing ^{3}H-BrdU was isolated by density gradient centrifugation, and its fragments—presumably containing origins—were cloned to form a library. The two longest fragments (pB48; 1560bp, and pLC46; 719bp) were used to make radioactive probes, which were then demonstrated to hybridize to early replicating genomic DNA. The pB48 fragment, being the longest fragment, was characterized using restriction mapping and Maxam-Gilbert sequencing (Tribioli et al., 1987). In a subsequent study, the genomic region represented by a 13.7kb insert containing the pB48 fragment was found to replicate within the first minute of S-phase (Biamonti et al., 1992). Fluorescence in situ hybridization mapped the region to the G-negative subtelomeric band p13.3 of chromosome 19, and the region was found to contain the 3'-terminus of human lamin B2 gene (Biamonti et al., 1992). Competitive PCR analysis—performed to quantify the abundance of BrdU-labeled nascent DNA—localized the origin to a ~500bp segment at the 3' noncoding end of the human lamin B2 gene (Giacca et al., 1997, 1994). The precise replication start site of the leading strands at the lamin B2 locus was determined at nucleotide resolution using replication initiation point (RIP) mapping (Abdurashidova et al., 2000). A 1.2kb DNA fragment containing the lamin B2 origin, when integrated at ectopic locations in the human genome, retained its

origin activity (Paixao et al., 2004). Together, these studies indicate that the human lamin B2 locus contains a site-specific origin that fires in early S-phase.

In addition to the human lamin B2 locus, site-specific origins have also been identified near the human beta-globin locus (Aladjem et al., 1998; Kitsberg et al., 1993) and the human c-myc locus (Waltz et al., 1996). This pattern of site-specific origin usage is in contrast to the origin distribution at the dihydrofolate reductase (DHFR) locus in Chinese hamster ovary cells (Hamlin et al., 2010). Origins were found to be in a broad zone—a 55kb region—at the DHFR locus, although controversies remain as to whether those origins are site-specific or whether they form a broad zone of inefficient origins. The latter model posits that there are many potential origins within the zone, and only a subset of them are activated in any cell cycle. As such, this distribution and usage of mammalian origins is in contrast to the site-specific, absolute efficient origin in *E. coli*. It has been posited that the site-specific nature of the human lamin B2 locus origin may be due to its geographic landscape: the locus is situated in a narrow intergenic region, which may physically restrict the pre-replication complex to initiate replication within a tight window (Hamlin et al., 2010).

While a handful of metazoan origins have been studied, many more potential origins are yet to be identified. If the average distance between origins is 100kb (Huberman and Riggs, 1968), then a 3 billion basepair human haploid genome would contain ~30,000 origins. Thus, to identify and understand all origins in the human genome, additional approaches are necessary.

1.1.8. Genome-wide mapping of replication origins

The difficulty in using the ARS plasmid assay to identify metazoan replicator sequences led to the development of other techniques. In general, the techniques fall into one of the following categories: 1) isolation of isotopically labeled nascent DNA; 2) identification of regions that have higher DNA content; 3) isolation of DNA bound to the initiator (ORC) or the replicative helicase (MCM); 4) isolation of newly synthesized DNA at the replication bubble; and 5) analysis of bubble-containing DNA structures.

After DNA isolation, putative origins can be identified by microarray or deep sequencing analyses. In *S. cerevisiae*, the ARS1 origin on Chromosome IV—the first origin to be identified (Stinchcomb et al., 1979)—became the prototype for investigating different origin-mapping techniques. Many of these genome-wide techniques have identified origins using cell-synchronization techniques that enrich for replication bubbles. In this section, I shall discuss the genome-wide techniques to map origins.

Isotopic labeling of nascent DNA

The first genome-wide origin mapping study adopted a Meselson-Stahl approach to isotopically label nascent DNA at the replication origins (Raghuraman et al., 2001). Budding yeast cells grown with ^{15}N were arrested in G1 phase and released into S-phase with medium containing ^{14}N to label nascent DNA. Density separation and microarray hybridization of the nascent DNA led to the detection of 332 putative origins. Several putative origins were isolated on restriction fragments and their origin activities were confirmed by the replication bubble structure on 2D gel electrophoresis.

Identification of origins by DNA content

Origins can also be detected by profiling DNA copy number (from 1N copy to 2N copy) on microarrays (Heichinger et al., 2006; Yabuki et al., 2002). When DNA replication begins at an origin, the DNA content of that locus increases from one copy to two copies. In a heterogeneous population of asynchronous cells, the copy number approach has limited utility because the DNA would be derived from all phases of the cell cycle. However, cell cycle synchronization techniques such as a G1 arrest using alpha-factor in *S. cerevisiae* (Yabuki et al., 2002) or a G2 arrest using the *cdc25-22* temperature-sensitive allele in *S. pombe* (Heichinger et al., 2006), followed by release into S-phase, have made the copy number approach informative. Additionally, the use of hydroxyurea has facilitated the identification and characterization of origins. Hydroxyurea inhibits ribonucleotide reductase, which is required for dNTP biosynthesis (Elford, 1968; Timson, 1975). Synchronized cells released into S-phase in the presence of hydroxyurea can initiate replication at a subset of origins ("early origins") before the replication forks stall (Santocanale and Diffley, 1998).

Chromatin-IP based approaches

The use of chromatin immunoprecipitation with microarray (ChIP-chip) or sequencing (ChIP-seq) has also provided insight into origin location. These approaches have been used to map the DNA location where components of the pre-replication complex, such as the origin recognition complex (ORC) initiator, and the minichromosome maintenance complexes (MCM) replicative helicase bind. Studies have been conducted in *S. cerevisiae* (Wyrick et al., 2001; Xu et al., 2006), *S. pombe* (Hayashi

et al., 2007), *Drosophila* (MacAlpine et al., 2010), and mammalian cells (Dellino et al., 2013; Miotto et al., 2016). A caveat of using ORC binding sites to infer the location of origins is that ORC has roles beyond DNA replication, such as transcriptional silencing (Bell, 2002).

Using ChIP-seq to map ORC1 binding sites in HeLa cells, ~13,600 ORC1 binding sites were mapped throughout the genome (Dellino et al., 2013). This study found an association between ORC1 binding sites and transcription start sites of coding and non-coding RNAs. Additionally, the transcription levels at ORC1 binding sites correlated with replication timing: origins that fire early in S-phase were associated with moderate-to-high transcription levels of coding RNAs, whereas origins that fire throughout S-phase were associated with low transcription levels of noncoding RNAs.

Using ChIP-seq to map ORC2 binding sites in asynchronous K562 human erythroid cells, ~52,000 ORC2 binding sites were mapped throughout the genome (Miotto et al., 2016). This study found that the ORC2 binding sites were not sequence specific. Rather, ORC2 binding sites were correlated with open chromatin regions (DNase I hypersensitive regions) that contain active histone modification marks. Other studies have also found that mammalian ORC seems to bind DNA with little sequence specificity (Remus et al., 2004; Schaarschmidt et al., 2004; Vashee et al., 2003). Together, these findings underscore the elusive nature of finding mammalian replicators, and suggest that the location of origins may be defined—to some degree—by steric or epigenetic factors. Additionally, these findings support a model whereby mammalian ORC binds indiscriminately to openly accessible regions to ensure replication is initiated.

Short nascent strand sequencing

Genome-wide identification of replication origins has also been done via mapping the location of short nascent DNA strands (i.e. newly synthesized DNA). At an origin, DNA replication initiation and elongation leads to the formation of short nascent strands (Vassilev and Johnson, 1989). The short nascent strands are eventually elongated to form the full length daughter strand, but if DNA replication elongation is interrupted (e.g. with the use of hydroxyurea to arrest replication), the newly synthesized DNA can be isolated and size-selected to yield a population of short nascent DNA strands (500bp to 1,500bp) that can be used for microarray or sequencing analyses (Gilbert, 2010). A limitation of this approach is that broken genomic DNA can contaminate the population. To enrich for newly synthesized DNA, investigators have performed BrdU pulse labeling in replicating cells followed by immunoprecipitation with anti-BrdU antibodies (Vassilev and Johnson, 1990, 1989; Vassilev et al., 1990).

Alternatively, enrichment of newly synthesized DNA can be achieved with lambda exonuclease, a 5’ to 3’ DNA exonuclease that digests DNA except for RNA-primed DNA. Lambda exonuclease thus effectively enriches for RNA-primed nascent DNA strands (Bielinsky and Gerbi, 1998; Gerbi and Bielinsky, 1997). Using lambda exonuclease enrichment, several groups have found that nascent strands were correlated with regions that potentially formed G-quadruplexes (G4 motifs), which are secondary DNA structures formed from guanine-rich motifs (Besnard et al., 2012; Cayrou et al., 2015, 2012; Mukhopadhyay et al., 2014; Picard et al., 2014; Valton et al., 2014). However, it remains controversial whether origins are actually correlated with predicted

G4 motifs, as lambda exonuclease activity is less effective at predicted G4 motifs (Foulk et al., 2015). Thus, the enrichment at predicted G4 motifs may be false positive artefacts of lambda exonuclease. Another study used BrdU-IP nascent strand enrichment without a lambda exonuclease step and found that origin efficiency was not correlated with the presence of predicted G4 motifs in a genome-wide, allele-specific study using human primary erythroblasts (Bartholdy et al., 2015). In this study, the origin profiles were compared between alleles that were sequenced and phased, and the differences in alleles due to single nucleotide polymorphisms, insertions, and deletions enabled a comparison of the effect of predicted G4 motifs on origin activity (Bartholdy et al., 2015). Thus, the biological significance of predicted G4 motifs to origin location and origin activity remains an active area of investigation.

A limitation of both the BrdU enrichment and lambda exonuclease enrichment approaches in mammalian cells is that the expected low origin efficiency leads to a low abundance of nascent strands, making it difficult to detect a signal. Another limitation of the BrdU-IP enrichment approach is the concern regarding epitope specificity.

Okazaki fragment mapping

Okazaki fragment mapping is another approach to identify origins. Okazaki fragments are short RNA-primed DNA fragments formed during the discontinuous synthesis of lagging strands. Okazaki fragments are strand-specific: they are generated either by a leftward moving replication fork (Watson strand) or a rightward moving replication fork (Crick strand). The strand-identity of an Okazaki fragment is thus a surrogate for replication fork direction. Due to bidirectional replication, a profile of the

Okazaki fragments around an origin should show a sharp transition in their strand placement. Using this approach in the context of DNA ligase I inactivation to enrich for Okazaki fragments, origins were found at sharp transitions in the distribution profile of the Watson and Crick strands of Okazaki fragments in *S. cerevisiae* (McGuffee et al., 2013; Smith and Whitehouse, 2012). The Okazaki fragment mapping approach has also been used in humans cells (Petryk et al., 2016), but the study found less sharp transitions that suggest broad zones of initiation sites (10-100kb each). This broad distribution draws similarity to that in the DHFR locus in Chinese hamster ovary cells (Hamlin et al., 2010).

Bubble trapping

The bubble trap method of identifying origins involves the isolation of DNA molecules that contain replication bubbles via gel electrophoresis (Eki et al., 2013; Mesner et al., 2013, 2011, 2006). This approach involves the restriction digestion of genomic DNA, which yields restriction fragments that contain either replication bubbles or no bubbles. The restriction fragments are subsequently separated via gel electrophoresis. The molten agarose, once solidified, effectively forms a mesh and interlocks bubble-containing DNA. The bubble-containing restriction fragments resemble closed circular structures that are "trapped" by agarose. Gel electrophoresis results in the separation of bubble-containing DNA (little or no mobility) from the non-bubble containing DNA (mobility). The bubble-containing fragments can then be cloned into plasmids and identified via microarray or deep sequencing.

1.1.9. Limitations of genome-wide ensemble approaches

The genome-wide origin mapping approaches to date have provided valuable insight into eukaryotic origins. However, these approaches average the behavior of a population of cells. Since eukaryotic origins do not fire in every cell cycle (that is, eukaryotic origins are not 100% efficient), bulk approaches suffer from signal-to-noise constraints. For example, an origin that fires with 5% efficiency (that is, an origin activated in 5 out of every 100 cell cycles) will—on average—produce a low abundance of nascent strands. The resulting signal may only be marginally above the background noise, making it difficult for low efficiency origins to be detected. Thus, the dynamic range at inefficient origins may be too low for signal detection.

Another limitation of bulk population approaches is that any information on the *cis*-acting effects of neighboring origins may be lost. For example, consider two neighboring origins shown in a hypothetical situation (**Figure 1.1**). A bulk population approach would show that both origins fire with 50% efficiency, but the two scenarios differ: the co-activation rate of the origins in the left scenario is 100%, whereas in the right scenario is 0%. The ability to determine how origin efficiency correlates with neighboring origin activity and origin density can provide insight into how eukaryotic DNA replication is coordinated to ensure the timely completion of genome duplication within S-phase.

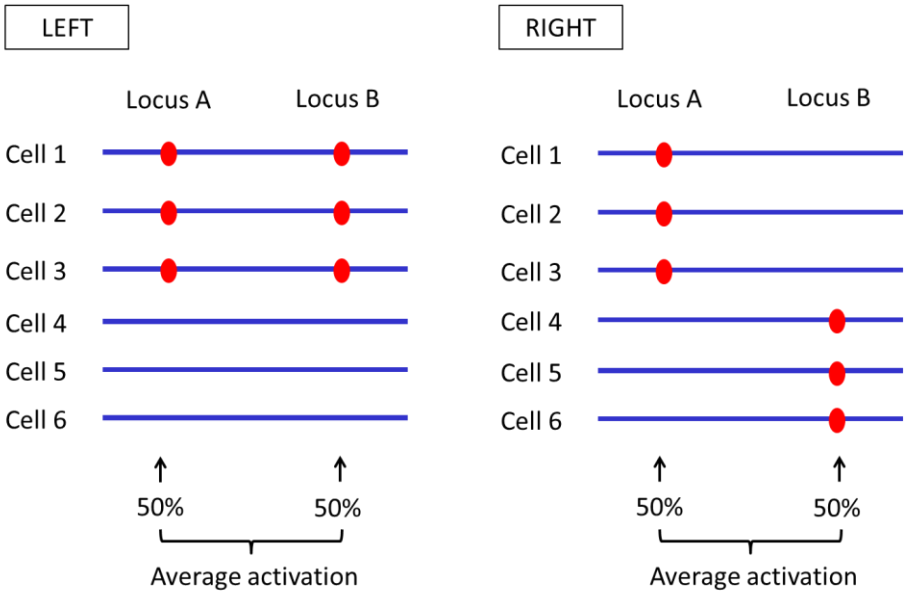


Figure 1.1. Single molecule approach can provide detailed origin efficiency information and nearby origin information.

Consider two hypothetical origins, one at Locus A and one at Locus B. Bulk analysis reveals the average origin efficiency at both Locus A and Locus B is identical (50%). However, a single molecule analysis reveals two different scenarios regarding the origin efficiency: On the LEFT, Locus A and Locus B co-activate at a frequency of 1.0, whereas on the RIGHT, Locus A and Locus B do not co-activate. Single molecule approaches can provide information on neighboring origins, which are unresolvable using bulk population averages.

1.1.10. Single-molecule mapping approaches

Since bulk population approaches provide limited resolution on the properties of origins with respect to its neighboring origins, it has become increasing important to use single molecule approaches to investigate origin behavior on individual DNA molecules. The first single molecule approach to study replication origins was done using ^{3}H-thymidine to pulse label sites of active replication, which were then visualized using DNA fiber autoradiography (Cairns, 1963; Huberman and Riggs, 1968). This method relies on radioactivity, which can be cumbersome and requires long exposure times to see the grain tracks (e.g. 4-6 month exposure in (Huberman and Riggs, 1968)). As a result, DNA autoradiography has largely been replaced by molecular combing (DNA combing) with immunofluorescence detection (Bensimon et al., 1994; Michalet et al., 1997).

In molecular combing, DNA fibers are stretched on silane-treated glass coverslips, which render the surface hydrophobic (Bensimon et al., 1994; Michalet et al., 1997). The process involves a mechanical device which dips a silanized glass coverslip into a buffered solution containing DNA for 5 minutes, after which the coverslip is vertically withdrawn from the solution at a constant speed. This process results in DNA molecules that are stretched and anchored on the glass coverslip, which can be visualized with microscopy (Bensimon et al., 1994; Michalet et al., 1997).

In molecular combing replication studies, thymidine analogs such as 5-bromo-2'-deoxyuridine (BrdU), 5-iodo-2'-deoxyuridine (IdU), and 5-chloro-2'-deoxyuridine (CldU) are used to pulse label cells, and analog incorporation tracks are visualized on stretched DNA fibers using fluorescent antibodies (Jackson and Pombo, 1998; Lengronne

et al., 2001). DNA combing can be combined with fluorescence in situ hybridization (FISH) to identify DNA molecules that hybridize to a specific probe (e.g. targeting a specific genomic location of interest) (Michalet et al., 1997). This approach has provided valuable information on the behavior of replication origins and replication forks. Indeed, studies on single DNA molecules—using DNA combing with FISH—indicate that origin firing in *S. cerevisiae* (Czajkowsky et al., 2008) and *S. pombe* (Patel et al., 2006) is stochastic, with each cell using a different cohort of origins from a large pool of putative origins (Bechhoefer and Rhind, 2012; Rhind, 2006). This feature of origin usage is consistent with the Jesuit Model ("Many are called but few are chosen") (DePamphilis, 1999) and is an effect of many origins being inefficient.

The Random Gap Problem

The variability in origin usage and origin efficiency raises the concern that some cells may not be able complete replication in time (Laskey, 1985). The random gap problem, as this is termed, describes the possibility that stochastic ("random") origin usage may lead to large distances ("gaps") between some adjacent origins that are unable to complete replication in time (Blow et al., 2001; Hyrien et al., 2003). Two models have been primarily proposed to address the random gap problem. The "regular spacing" or "fixed spacing" model (Hyrien et al., 2003) posits that while origins may be randomly distributed, the spacing between origins are regular and short enough to ensure replication completion (Blow et al., 2001; Hyrien and Mechali, 1993). A second model, the "increasing origin efficiency" or "origin redundancy" model (Hyrien et al., 2003) posits that an "excess of potential origins" are licensed in G1 phase, and that these

licensed origins have increasing probabilities of firing as S-phase progresses. (Eshaghi et al., 2007; Herrick et al., 2000; Lucas et al., 2000; Rhind, 2006; Yang and Bechhoefer, 2008). Thus, any unfired putative origin in large gaps have a higher propensity to fire as S-phase progresses. The "regular spacing" model appears to have "little experimental support" (Jun and Rhind, 2008), whereas the increasing origin efficiency model appears more favored (Hyrien et al., 2003; Jun and Rhind, 2008). A third model—which has yet to be widely accepted—posits that replication fork velocities are dynamically regulated, with replication forks progressing at higher velocities when the interorigin distances are larger (Conti et al., 2007). This model is based on single molecule DNA combing observations from pulse labeling experiments with two color fluorescence (two different nucleotide analogs). This approach entails pulse labeling replicating cells with two sequential, independent pulses of thymidine analogs (e.g. IdU then CldU) for known durations. Analog incorporation tracks are visualized using fluorescence antibodies (Jackson and Pombo, 1998), and their incorporation patch lengths can be measured to determine the replication fork velocity. Although this approach is valid for determining replication fork velocities, any variations in DNA stretching or any non-specific attachment of fluorescence antibodies to the DNA molecule—which do occur in DNA combing—can lead to variations in the calculated replication fork velocities. Additionally, the halogenated nucleoside analogs need to be metabolized to the triphosphate form and compete with the unlabeled endogenous nucleotide pool for incorporation. The potential imprecisions of using DNA combing to calculate replication fork velocities have provided motivation to develop single molecule approaches with higher resolution and

higher throughput to determine the replication fork rate, which shall be explored in Chapter II and Chapter III. Indeed, a bulk approach using ChIP against the GINS complex—a component of the replication fork—followed by microarray suggests that replication fork speed is uniform throughout the *S. cerevisiae* genome (Sekedat et al., 2010).

A third single molecule approach to studying DNA replication is electron microscopy. This technique can provide topological information, such as the presence of replication bubbles and replication forks on a chromosome, and/or the presence of protein-DNA complexes such as the MCM2-7 complex loaded on DNA (Evrin et al., 2009; Petes and Newlon, 1974). Unlike molecular combing, where the DNA molecules are stretched into linear forms, electron microscopy can resolve DNA knots and replication intermediates such as Holliday junctions (Sogo, 2002; Sogo et al., 1999). As such, electron microscopy can also be useful for studying DNA repair and DNA recombination. The topological appearance, however, is not specific for replication intermediates, because multiple superimposed DNA molecules can result in a similar appearance. Importantly, electron microscopy does not provide sequence information or the genomic location of origins, and it has not been adapted for mapping origins.

Similarly, single molecule approaches such as colocalization single-molecule spectroscopy (CoSMOS) with fluorescence resonance energy transfer (FRET); real-time visualization of protein-DNA interactions with fluorescence microscopy or total internal reflection fluorescence (TIRF) microscopy; or DNA curtain combined with TIRF, have been used to investigate pre-recognition complex assembly and replisome formation

(Duzdevich et al., 2015; Graham et al., 2017; Ticau et al., 2017, 2015; Yardimci et al., 2012). These are not mapping techniques designed to identify origins on a genome-wide scale, but are aimed at understanding the molecular mechanisms of origin activation and origin licensing. The current state-of-the-art approach to mapping origins at a genome-wide scale is thus limited to DNA fiber analysis (molecular combing with FISH).

1.1.11. Limitations of current single molecule mapping approaches

The development of molecular combing techniques to study replication origins and replication kinetics is a significant improvement to the DNA fiber autoradiographic studies that require long exposure times. However, there are several challenges to employing molecular combing to study replication origins on a genome-wide scale. First, although molecular combing provides an optical view of individual DNA molecules, it does not provide DNA sequence or the genomic source of individual DNA molecules. When combined with fluorescence in situ hybridization (FISH), molecular combing can be used to identify DNA molecules from a specific site (Michalet et al., 1997). However, using this method to map DNA molecules across the genome would require many different FISH probes, with each probe targeting an independent region of the genome. Using this technique on a genome-wide scale is low-throughput. Second, the spatial resolution of molecular combing is limited by optical detection of fluorophores. Molecular combing does not provide the DNA sequence information, and the resolution is not at a single-base resolution. Variations in antibody fluorescence staining or DNA molecule stretching can affect the calculated replication fork patch length (and hence fork speed). These limitations call for the development of new single molecule approaches to study replication origins and replication kinetics. The need for new single molecule approaches is particularly true for investigations aimed at identifying inefficient origins, and for studies aimed at addressing questions such as the random gap problem.

1.2. Rationale for thesis

Since the replicon model of DNA replication was proposed in 1963, scientists have searched for origins to gain a better understanding of DNA replication (Jacob et al., 1963). The *E. coli* origin, being the only replicator sequence in its circular chromosome genome, is well-defined at a precise location (Blattner et al., 1997). Eukaryotes, however, have more complex genomes with multiple chromosomes and multiple origins (Huberman and Riggs, 1968). Additionally, eukaryotic origins are not all activated simultaneously or activated in every cell cycle. Origins are a critical point of regulation of replication. To appreciate how DNA replication initiation is regulated, it is important to understand where and when DNA replication starts. Genome-wide mapping techniques have been used to identify putative origins in higher eukaryotes. These studies have advanced our understanding of the origin location and replication timing of a bulk population of cells. However, bulk approaches average the behavior of a large population of cells, and cannot resolve cell-to-cell variations in origin activation or replication fork progression.

To provide higher resolution in identifying origins and determining their efficiency of activation, we have utilized three emerging technologies to map origins on single DNA molecules. The Pacific Biosciences (PacBio) Single Molecule Real-Time (SMRT) sequencing platform is a parallel sequencing technology capable of generating long reads (N50 >30kb) without template amplification. Recent studies have identified modified bases such as 5-methylcytosine and N6-methyladenine using SMRT sequencing

(Flusberg et al., 2010). In Chapter II, I postulate that SMRT sequencing can be used to detect thymidine analogs, such as BrdU, EdU, and IdU. These analogs can be pulse-labeled into cells to mark the replication start site. Pulse labeled DNA can subsequently be SMRT-sequenced to map sites of replication. In Chapter III, I explore the feasibility of using the BioNano Genomics Irys platform to optically map sites of replication. The Irys system is a high-throughput genome mapping technology that can automatically image many parallel DNA molecules in each camera frame. In Chapter IV, I briefly discuss the prospect of using nanopore sequencing to map sites of DNA replication. Nanopore sequencing is a DNA sequencing technology that is rapidly evolving. The technology relies on the passage of DNA molecules through narrow pores, which changes the electrical conductance across the nanopore in a sequence-dependent manner. Thus, the DNA sequence can be determined by the signature electrical conductance changes. Based on this principle, we posit that modified nucleoside analogs in pulsed labeled DNA, such as BrdU, EdU, and IdU, can be nanopore-sequenced and used to identify replication sites.

Chapter II: Detection of thymidine analogs using Single Molecule Real-Time sequencing

2.1. Introduction

DNA replication is a key biological process whose precision and accuracy in duplicating the genome can affect the daughter cell genotype and fate. *E. coli* has one well-defined origin of replication that occurs at a specific site in the genome (Bonhoeffer and Gierer, 1963). In higher eukaryotes, the location and behavior of origins are not well-understood. Genome-wide approaches to map origins—such as nascent strand abundance mapping, Okazaki fragment mapping, or BrdU-IP combined with microarray or deep sequencing—average the behavior of a population of cells. Due to cell-to-cell variability in replication origin usage, single molecule approaches are necessary to investigate the actual behavior of a cell. Single molecules approaches available to date—including DNA fiber autoradiography or molecular combing in conjunction with fluorescence in situ hybridization (FISH)—are low-throughput techniques that provide a visualization of the replication tracks. The ability to study replication kinetics on single DNA molecules at high-throughput will provide a greater understanding of DNA replication.

Single molecule real-time (SMRT) sequencing, developed by Pacific Biosciences (PacBio), is capable of providing the sequence of long DNA molecules (Eid et al., 2009). Our experiments with the PacBio RS II instrument for SMRT sequencing have all yielded N50 values >30kb. The technology involves: 1) a DNA polymerase to sequence the template DNA; and 2) fluorescent phospholinked nucleotides, with each of the four canonical bases, A, C, G, and T, being represented by a different colored fluorophore (Korlach et al., 2010). Incorporation of a nucleotide cleaves the phosphate linker and

releases its fluorescent tag, which can be optically captured to determine the base identity (**Figure 2.1**). The DNA polymerase is anchored to the bottom of nanostructures called zero-mode waveguides (ZMWs). Each chip (SMRT cell) contains 163,482 ZMWs, which serve as visualization chambers to capture the fluorescence emitted during the sequencing reactions. The fluorescence intensity traces are then processed via base-calling algorithms to determine the template DNA sequence.

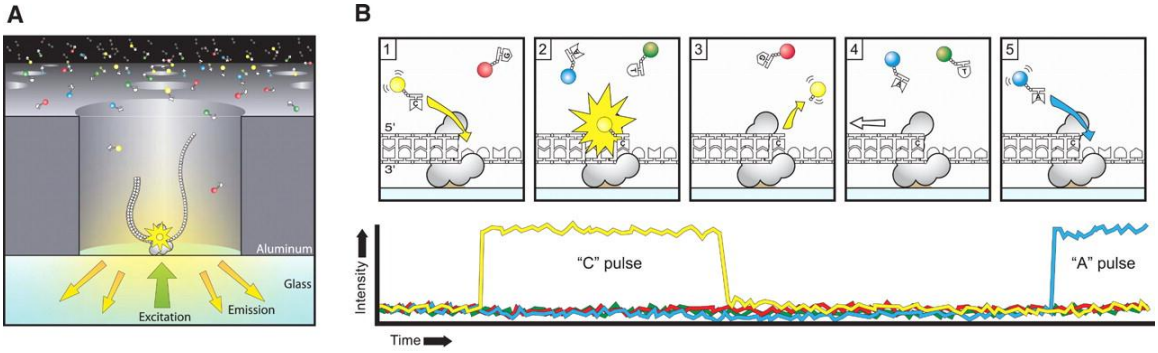


Figure 2.1. Single molecule real-time (SMRT) sequencing.

A) A DNA polymerase—attached to the bottom of nanostructures called zero-mode waveguides (ZMWs)—is used to sequence the DNA template. Incorporation of fluorescent phospholinked nucleotides releases the fluorophore, whose intensity can be recorded in each ZMW. B) The fluorescence is distinct for each of the four canonical bases. The fluorescence intensity pulse is used to generate the template DNA sequence via base calling algorithms. From Eid J. et al., Real-time DNA sequencing from single polymerase molecules. Science, 2009, 323: 133–138. Reprinted with permission from AAAS.

The standard DNA library preparation for SMRT sequencing involves: 1) isolation of the double-stranded DNA template; 2) end-repair of the template; 3) ligation of hairpins to both ends of the template. The resulting DNA structure, termed SMRTbell for its dumbbell shape, serves as the template for sequencing (Travers et al., 2010). SMRTbells are loaded onto a chip ("SMRT cell") that contains 163,482 ZMWs. A primer that binds to the hairpin region is used, along with fluorescent phospholinked nucleotides and a DNA polymerase with strand displacement activity, such as phi29. The strand displacement activity allows the DNA polymerase to undergo rolling circle amplification, such that each template can be sequenced with multiple passes (**Figure 2.2**) (Travers et al., 2010).

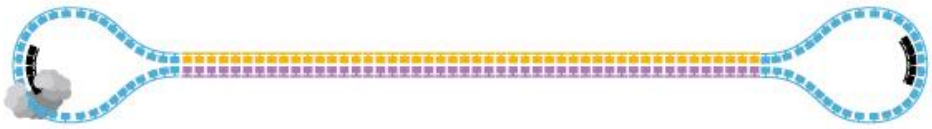


Figure 2.2. SMRTbell template for sequencing.

Double-stranded DNA are processed and ligated with hairpins (blue) to form a dumbbell-like structure. Primers (black) are designed to anneal to the hairpin region. Figure reproduced with permission from Pacific Biosciences.

The fluorescence traces captured in each ZMW can be processed via base calling algorithms to produce the nucleotide sequence of the DNA template. Several types of reads—each containing a DNA sequence—can be generated from the standard PacBio analysis workflow on SMRTbell templates (**Figure 2.3**). The raw read is the entire base-called DNA sequence generated from a ZMW. Ideally, the raw read represents all incorporation events along the template, which consists of the insert DNA and the hairpin adapters. However, because raw reads are unfiltered, any fluorescence—signal or noise—captured in the ZMW can affect the sequence. The polymerase read is the DNA sequence after trimming out low quality regions, as determined by the standard PacBio algorithm. Due to rolling circle amplification, an unfiltered raw read and a polymerase read can contain multiple repeats of the DNA template. The unit of DNA sequence that is repeated is called a subread. Subreads ideally represent the insert DNA and exclude the hairpin adapter sequences. Multiple subreads can be aligned to establish the consensus sequence at each position of the insert DNA. This sequence is the circular consensus sequencing read (ccsRead).

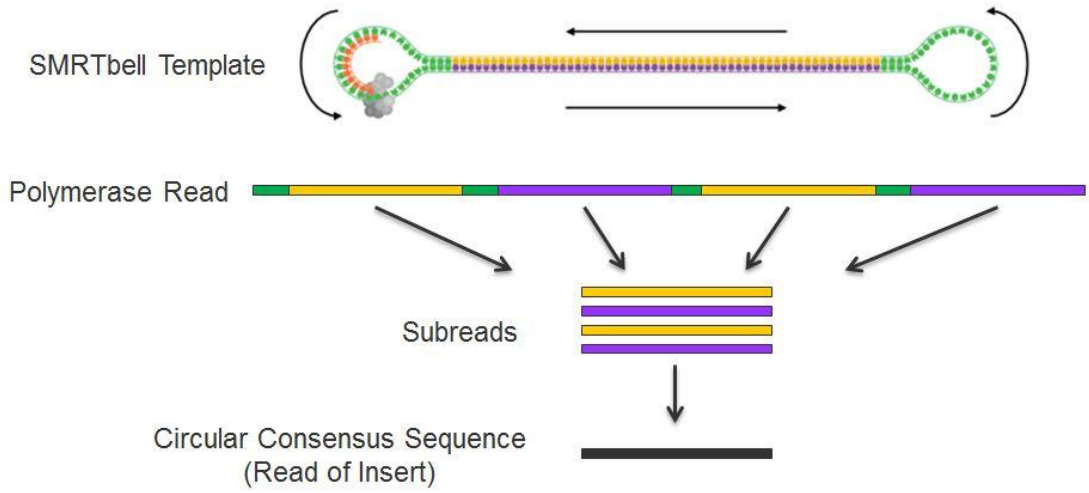


Figure 2.3. Types of reads generated from SMRTbell template sequencing.

The raw read (not shown) is the unfiltered, complete base-called DNA sequence generated in a ZMW. The polymerase read is the DNA sequence after trimming out low quality regions from the raw read. A subread is the unit of DNA sequence that is repeated in multiple rounds of sequencing. Multiple subreads can be used to generate a circular consensus sequence (ccsRead). Figure reproduced with permission from Pacific Biosciences.

Among the applications for SMRT sequencing—which include *de novo* genome assembly and structural variation detection (Rhoads and Au, 2015)—the most pertinent application to our replication kinetics studies is base modification detection. Modified bases, such as 5-methylcytosine and N6-methyladenine, can be detected in DNA templates using kinetic measurements of nucleotide incorporation (Flusberg et al., 2010). As the DNA polymerase travels along the template, each nucleotide incorporation event releases a fluorophore whose fluorescence intensity pulse is recorded over time. Two kinetic parameters can be derived from the fluorescence intensity pulse. The pulse width (PW) is the duration of a fluorescence pulse, and reflects the time it takes a polymerase-bound nucleotide to be incorporated. The interpulse duration (IPD) is the time elapsed between the incorporation of a nucleotide and its previous nucleotide (**Figure 2.4**). Studies have shown that in synthetic DNA templates containing 5-methylcytosine, 5-hydroxymethylcytosine, or N6-methyladenine, the IPD is significantly higher for the modified base compared to their respective unmodified base control (Flusberg et al., 2010; Song et al., 2011). Using this property of sequencing kinetics, investigators have used SMRT-sequencing to identify N6-methyladenine in *E. coli* (Fang et al., 2012) and in mammalian cells (Wu et al., 2016). Other modified bases have been successfully detected via SMRT-sequencing, including 8-oxoguanine, 8-oxoadenine, O6-methyl-guanine, 1-methyladenine, O4-methylthymidine, 5-hydroxycytosine, 5-hydroxyuracil, 5-hydroxymethyluracil, thymine dimers, and

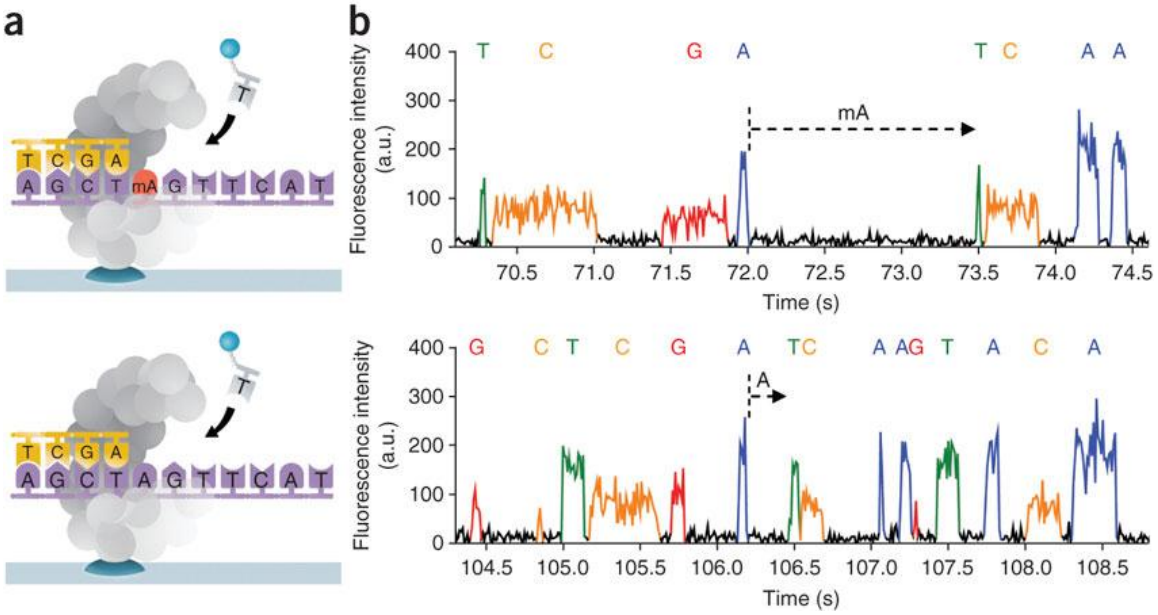


Figure 2.4. Detection of modified bases via SMRT sequencing using the IPD.

a) Schematic depicting polymerase incorporation along a DNA template containing methyl-adenine (mA) and a DNA template containing unmodified adenine (A). **b)** The interpulse duration (IPD) is represented by the horizontal dash line. In this example, the IPD is longer for methyl-adenine (mA) in the template compared to unmodified adenine (A). Reprinted by permission from Macmillan Publishers Ltd: Nat. Methods 7: 461–465, copyright 2010 (Flusberg et al., 2010).

beta-D-glucosyl-hydroxymethyluracil (Clark et al., 2011; Genest et al., 2015; Korlach and Turner, 2012). Based on these findings, I hypothesized that thymidine analogs—substrates frequently used in replication kinetics studies—can be kinetically detected via SMRT-sequencing. Furthermore, I postulate that the detection of thymidine analogs can be used to map the locations of active DNA replication on single DNA molecules.

SMRT-sequencing can advance existing technology for mapping origins

Halogenated thymidine analogs such as 5-bromo-2'-deoxyuridine (BrdU), 5-chloro-2'-deoxyuridine (CldU), and 5-iodo-2'-deoxyuridine (IdU), have been used extensively in pulse labeling experiments to mark the sites of active DNA replication. BrdU-labeled DNA can be used to map origins using either bulk approaches or single molecule approaches. However, existing approaches have drawbacks that can be addressed by SMRT-sequencing.

First, bulk approaches such as BrdU-immunoprecipitation followed by microarray or Illumina deep sequencing can provide a genome-wide profile of replication origins, but these approaches average the enrichment of BrdU-labeled DNA over control DNA from a population of cells. Consequently, these approaches are difficult to detect replication origins that fire inefficiently (that is, regions where the origin usage is low). At inefficient origins—regions with low frequency of replication initiation—the enrichment of BrdU-labeled DNA is expected to be low. Thus, these bulk approaches have a low dynamic range for identifying inefficient origins. In contrast to bulk approaches, SMRT-sequencing analysis is performed on independent single DNA

molecules. As in the case of 5-methylcytosine versus cytosine (Flusberg et al., 2010), we expect a high discriminatory power for the IPD: the modified base should provide a significantly higher IPD than its unmodified counterpart. Thus, even at inefficient origins—for example, where only 5% of cells initiate replication at a particular genomic locus—we still expect the modified base to be detected among the 5% of individual DNA molecules that are SMRT-sequenced. The ability to identify inefficient origins and to determine replication fork velocities in single DNA molecules on a genome-wide scale would advance our understanding of DNA replication.

Second, Illumina deep sequencing utilizes short reads and requires template amplification, which can introduce coverage biases. In contrast, SMRT-sequencing utilizes long reads and does not require template amplification.

Third, antibodies introduce a concern regarding epitope specificity and accessibility to the incorporated halogenated thymidine analog. SMRT-sequencing does not require antibodies.

Fourth, existing single molecule approaches to detect replication origins and replication forks—DNA fiber autoradiography and molecular combing (DNA combing)—have limited spatial resolution. Both approaches can provide a visualization of replication tracks, either as grain tracks or fluorescence signal on DNA fibers (Bianco et al., 2012; Cairns, 1963; Huberman and Riggs, 1968; Jackson and Pombo, 1998; Pasero et al., 2002). However, these existing approaches fall short in comparison to SMRT-

sequencing, which can provide high resolution (single base level resolution) nucleotide sequence readouts of individual DNA molecules.

Based on SMRT-sequencing kinetic studies of modified bases, I hypothesized that halogenated thymidine analogs can be detected via SMRT-sequencing. Here, I tested the ability to detect high IPD values in PCR products containing BrdU, IdU, or the non-halogenated thymidine analog 5-ethynyl-2'deoxyuridine (EdU). The results suggest that thymidine analogs can be detected via SMRT-sequencing. This proof-of-principle finding suggests that SMRT-sequencing can be used to study replication origins and replication forks in organisms from *E. coli* to yeast to mammals.

2.2. Materials and Methods

Oligonucleotide and plasmid sequences are tabulated in **Tables 2.1** to **2.3**.

Reagents

Modified deoxyribonucleoside triphosphates were purchased commercially:

5-bromo-2'-deoxyuridine-5'-triphosphate (BrdUTP) from Sigma Aldrich; 5'-ethynyl-2'-deoxyuridine-5'-triphosphate (EdUTP) from Jena Bioscience; and 5-iodo-2'-deoxyuridine-5'-triphosphate (IdUTP) from TriLink Biotechnologies, San Diego, CA, USA. Canonical dNTPs (dATP, dCTP, dGTP, dTTP) were purchased from New England Biolabs.

Template for PCR reactions

The plasmid pFS474 contains a 1,312bp insert (part of a 6-mer de Bruijn sequence) cloned into pUC57, and was custom synthesized (GenScript, Piscataway, NJ). For libraries containing modified bases in the region of interest, primers were designed to target a 1.6kb region spanning the insert in pFS474 to produce a 1.8kb amplicon (1.6kb target plus two ~100-mers). For the single thymidine-to-EdU modification library, custom oligonucleotide with one modified base was purchased from IDT, Integrated DNA Technologies Inc., CA.

Table 2.1. Oligonucleotide sequences for the preparation of thymidine and thymidine analog-containing PCR products.

Product ID	Library	Primer ID	Primer sequence (5' → 3'). Uppercase denotes barcode sequence. Lowercase denotes priming sequence that binds to pFS474 plasmid.
VLPB001	BrdU	VL-318	TTGGTGTCTTCTCGTGCTCGCCTGTGCTGTGTGT GCTGCTGTCTGCTCTGTGCCGTGTCTTCCGTCTG CGTGGCGCCGCTTGTCGTggtgcgggcctcttcgctat
VLPB001	BrdU	VL-325	TGCGTCTCGTTTTGTGTCTTTCCGTGCTGCGTGG CTCTTGCGCGCCTTGTCGTGGTCTCTTCGTCTGC TGCTCTCgtttcccgactggaaagcgggcagt
VLPB002	EdU	VL-319	TTGGTGGTCTCGTCTGCTCGTGTCGTGTCTCGCT GTCTGTGCCTGTGCGTGGTCTTCCGTGTGTCTGT GTCCTGCTGTTCTTGGCGgtgcgggcctcttcgctat
VLPB002	EdU	VL-326	TGCGTCCGCTGTTCGGTGTTCTCTCTGTTCCTGG TGTGTTCTCTCTGCCGGTTCTCGCGTCTGGTTTG GTCCTCTgtttcccgactggaaagcgggcagt
VLPB003	IdU	VL-320	TTGGTGTTCGCCTCGTCTGTTCGTCCTCTGCTCG GTGTCTGCGTCGTGTGCGGCTTTTGCTCCTGCTC GTTGTGTTCTCCGTCGCTggtgcgggcctcttcgctat
VLPB003	IdU	VL-327	TGCGTCTGTTGGGTGTGCTTGCTTCTCCTGTTGC TGTGTTGTGCGTTTGTGCGGTCTCTGCTCGTGTC GTTCTGTgtttcccgactggaaagcgggcagt
VLPB004	Mixed BrdU & T	VL-317	TTGGTGCTCGGTTGGCCTCTCGGTTGCTCGCGC GTGTTCGTTGCCGGTGCGTCTCTGCGTGTGCCCC TCCTGCTCTTCGTTGTGCTggtgcgggcctcttcgctat
VLPB004	Mixed BrdU & T	VL-324	TGCGTCTTGTGTGTGTGCGCGTCTGGGTGCTGT GCTGCTCGTCTCCTGGTGGTCCGCTTGTCCGTTT CTCCTGCCgtttcccgactggaaagcgggcagt

Table 2.1 (continued). Oligonucleotide sequences for the preparation of thymidine and thymidine analog-containing PCR products.

Product ID	Library	Primer ID	Primer sequence (5' → 3'). Uppercase denotes barcode sequence. Lowercase denotes priming sequence that binds to pFS474 plasmid.
VLPB005	Mixed EdU & T	VL-322	TTGGTGGGTGCCCTGTGCTGCCTTGTCTTGGCGT GTCCGTCTGTGGCTTTGTGTCTCTTGCGTCTTCC TTCGTTGTCGTGTCTCGGggtgcgggcctcttcgctat
VLPB005	Mixed EdU & T	VL-329	TGCGTCCTGTCCTCGTGGTGGTCTCGTGCGTGCT TGTCTCTTTCGTGCGTCTTCTGTGCTGTTCGTGG TGTGTCCgtttcccgactggaaagcgggcagt
VLPB006	Mixed IdU & T	VL-302	TGTACTTTGTTCGCTGTGGGCTGCTGTCTCCTCT CGGCTCTCTCTGCGGTGTTCTTGCTGTGCGCTTC TCggtgcgggcctcttcgctat
VLPB006	Mixed IdU & T	VL-311	TGCGTCATGATCCTCGTCTCCCTCTCCGTTGTGT CCTCGTCTCTCCTGTCTCTCTGGTTTCCGGTTCG TGTGTGGgtttcccgactggaaagcgggcagt
VLPB007	T	VL-303	TGTACTTGCTCTTCGTCCTGCCTCTGCGTGTCGT GTGTGTCGCCCTTCGTCTGCTCGTCTCTGCTTTG TTggtgcgggcctcttcgctat
VLPB007	T	VL-312	TGCGTCATGATGGTGTGCTCTTTTTTGGGCTGCG TGCGGCTGGTTCCGTTGGTTTGGGTCCGTTCGCC TCTCTTGgtttcccgactggaaagcgggcagt

Table 2.2. Oligonucleotide sequences for the preparation of control template (thymidine only) and single thymidine-to-EdU substituted PCR product.

Product ID	Library	Primer ID	Primer sequence (5' → 3'). Uppercase denotes barcode sequence. Lowercase denotes priming sequence that binds to pFS474 plasmid.
VLPB015	T	VL-337	GACTCTATGAGCACAGAGTCTTTATCACGCACT GATCAGCCAAATCATCCGCCAGTATTGCTGTGA CGGAggtgcgggcctcttcgctat
VLPB015	T	VL-372	TTATCTCCCGCAACCGTCTGTGACGAGCTCTGG ACACACTGCTGGATTTCTCATGATGCGATGATA GAACattaaagtatttatggcggc
VLPB016	T (with 1 EdU)	VL-375	GACTCTATGAGCACAGAGTCTTTATCACGCACT GATCAGCCAAA[EdU]CATCCGCCAGTATTGCTG TGACGGAggtgcgggcctcttcgctat
VLPB016	T (with 1 EdU)	VL-373	TCTCTATTGATCATCTCGCGACTGTAGTAGAGA TTGCCTGATGAGTGCTCTACTCTATGCTACACTC GTGattaaagtatttatggcggc

Table 2.3. Plasmid pFS474 sequence.

DNA sequence is 5’ to 3’. The primer binding sites are underlined.

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGA
TGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATC
AGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAG
GCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCG
AAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCC
AGTGAATTCGAGCTCGGTACCTCGCGAATGCATCTAGATGGGGGTGCACGCTTATACGCTCTGGAGGCCCGAGGTAAT
AATAAATTAATAAAATAAGGTAATCGAATCTCAGGGCGACCATGAAGTTGTGTAGAGATATAGGAAGGATGCAACACG
CCTCTGAAAATTTAGCCCGCACTGCTCATGATCTAACAGTCTTGCTAGATTTTAAGGGCCGTCCCTCGCTCAGTGAAAC
ACCAACAAAGAACAATAGAACCGTATTAATATATTACATTATTACCGTAGCAATCCACGGACCGCGGGTACGCACAGG
AACAGACCCAGAATACCGAATGAGTTGGACTTGGCCGCCATAAATACTTTAATACATAATGCCATAACCTGGGTTATTG
GGTATATTTATTTATTTATATATGGGTATCTGGGTGAAGCCTTAAAGCCACAGCTCCGGGCATATAGGCATGTAGTAGC
CTTTGACCCGCGTAATTAAATCTATAAGTAGAACATCTACCAAAAGCACCGAGTCCCAATTTTCGGGCGTTGCAATTAG
CGCGATCAGGACTAGCGTAGATCCAGATCTGTGCCCGGAAGACTAATTTCAATCGTAGTTCAGGCGCACGGTTATTATT
TTATTCTTATATTCGGTTGGTTATTTATATATTTATTTATTTGGTAGGCCAGGTCTAATGTAATAATGTAATAATGGTCTT
TAACGAATAGGCGGTCGTATTCAGTCCGACTATCCTTTTGTAACTTCACATGGCGTCAACGTTACACGTGCGATAAGGC
CGGGTGTCGCGTGAACCATTAGTGCAAATGTGCAGCCCCACCATCCGGACTGGTGAGTGGTAAAATTGATTAATTGTAA
AGTGGCGCTTTCTTTGTGGCCCACATCAATTCCCTATATCATAATATCATAATATCCCTAACCCAACTCTGCGTACCTAC
TATCATTCTATCATTCTATCCTACAGAAAAAGGACCCTAGAAACTATGATTCTATGAAACCTAGAAGATAACATAATAA
CGCCACCTCTCACCTTGCAGGGTCGCTTCAAGGGAGACCAGTAAACGGCAGACAGAGTAGTCTAGGGCTTATTTAAAA
CAAGGAAACTCGCCCATATCTCCCCTCATTGGGATAATTGCGCAGAACGTAACCGCCCTCCCAGTTATACGTTGAACGG
ATTGATAAATAAACCTAGGACAAATGAATATCCGTTCTTAGATAGAGGCTGTGTCATTCGTGAGCCGCAACGGTCCGGC
AAACCAGCAGCTAGCAGAGAAAGCGCTCGCATATAATGGAAAGGGTAAGGAGTCTTATAGTCTGCCTCGAAGGCTCAC
GGGATCCAATATCGGATCCCGGGCCCGTCGACTGCAGAGGCCTGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTT
TCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCC
TAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGC
ATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCT
GCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGA
TAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTT
TCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTAT
AAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCC
GCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTC
CAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAAC
CCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCT
ACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAG
TTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAA
GCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAA
CGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGA
AGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCT
CAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCA
TCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCC
GGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAG
TAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGT
ATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCT
CCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCT
CTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG
GCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATT
GGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCAC
CCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAA
GGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATT
GTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGT
GCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTC

Barcode design

Barcodes composed of only three canonical bases, C, G, and T, were designed to ensure that the barcode and its complement do not contain thymidine analogs. This design facilitates the identification of the PCR product. Unique barcodes are located at the 5' ends of each primer. The barcode sequences were provided by John Urban (Brown University). Barcode pairs (each 75-86 bases in length) were manually screened for lower probabilities for homodimerization and heterodimerization. To decrease the risk of off-target priming, the terminal 6-mer at the 5'-end of each barcode was screened to ensure its complement does not bind the plasmid template.

Preparation of PCR products with a 1.6kb region that contain thymidine analogs

PCR was performed using barcoded primers that bind to sites flanking the insert in pFS474. PCR was performed in 50uL reactions using the following reagents: 1ng plasmid pFS474, 62.5nM forward primer, 62.5nM reverse primer, 1x Pwo SuperYield Buffer with Mg^{2+}, 0.2mM of each deoxyribonucleoside triphosphate (dATP, dCTP, dGTP, dXTP, where dX = BrdU, EdU, IdU, T, or a 1:1 mixture of the following pairs: BrdU & T, EdU & T, or IdU & T), and Pwo DNA polymerase (Roche).

Touchdown PCR was performed with the cycling conditions: 95^{o}C for 3mins, then 30 cycles (95^{o}C for 30secs, 65^{o}C for 45secs (-0.5^{o}C per cycle), 72^{o}C for 2mins), then 5 cycles (95^{o}C for 30secs, 50^{o}C for 45secs, 72^{o}C for 2mins), then 72^{o}C for 5mins, then 4^{o}C for 20mins.

Preparation of PCR product with single thymidine-to-EdU substitution

A custom oligonucleotide with a single EdU at position 45 was purchased (IDT, Integrated DNA Technologies Inc., CA). This oligonucleotide (VL-375) consists of only 90 nucleotides, with all bases (except for EdU at position 45) being either A, C, G, or T. To place the single EdU in a larger DNA template to facilitate library preparation and sequencing, this oligonucleotide was used as a forward primer in a PCR approach. For control, an oligonucleotide (VL-337) with the identical sequence but with thymidine at position 45 was purchased (Sigma Aldrich). Two oligonucleotides (VL-373 and VL-372 respectively) with only canonical bases were used as reverse primers and have barcode sequences to facilitate analysis. PCR was performed using canonical dNTPs (dATP, dCTP, dGTP, dTTP), such that the full-length experimental PCR product (627bp) has only one modified base: EdU at position 45.

Touchdown PCR was performed with the cycling conditions: 95^{o}C for 3mins, then 30 cycles (95^{o}C for 30secs, 65^{o}C for 45secs (-0.5^{o}C per cycle), 72^{o}C for 30secs), then 5 cycles (95^{o}C for 30secs, 50^{o}C for 45secs, 72^{o}C for 3min), then 72^{o}C for 5mins, then 4^{o}C for 20mins.

Library preparation

PCR products were cleaned up with 0.8x volume of Agencourt AMPure XP beads (Beckman Coulter, Inc.) according to the manufacturer's protocol with a minor modification. At the recommendation of PacBio staff (personal communication), AMPure XP beads were pre-washed by first immobilizing the beads, saving the original buffer, rinsing the beads with water for multiple times, and then placing the original buffer back into the beads. After purification, PCR products were eluted in 1x TE and sent to the Deep Sequencing Core at UMass Medical School for library preparation and PacBio SMRT sequencing. Library preparation was performed according to standard PacBio protocol, and included end-repair of the DNA templates and ligation of adapters to form SMRTbells. Samples were multiplexed and SMRT sequenced using the PacBio RS II instrument with P6C4 chemistry and 360 minute data collection time. Fluorescence traces were recorded at 75 frames per second.

Alignment of reads to template

Raw data in bax.h5 files were analyzed using SMRT Analysis (v2.3.0). To align filtered reads to the reference template, the PacBio alignment utility, *pbalign*, was used with the options *--forQuiver –metrics IPD,DeletionQV,DeletionTag,InsertionQV,MergeQV,SubstitutionQV*.

To assign barcodes to the aligned reads, the PacBio barcoding utility, *pbbarcode*, was used to first label ZMWs and then to label aligned reads. Aligned reads annotated with their respective barcodes were then separated into groups based on their barcodes using the PacBio script *cmph5tools.py*. Next, polymerase kinetics, including the IPD, were loaded using the PacBio script *ipdSummary.py*.

IPD ratio and mean of the IPD ratio

The IPD ratio can be calculated after aligning reads to the reference template. The IPD ratio is equal to the average IPD at a particular position in the sample library, divided by the average IPD at that position in the control (no analog) library. The IPD ratio at all positions on the template can be averaged to produce the mean of the IPD ratio.

2.3. Results

To determine whether thymidine analogs in PCR products are detectable by SMRT-sequencing, I took an approach outlined in **Figure 2.5**.

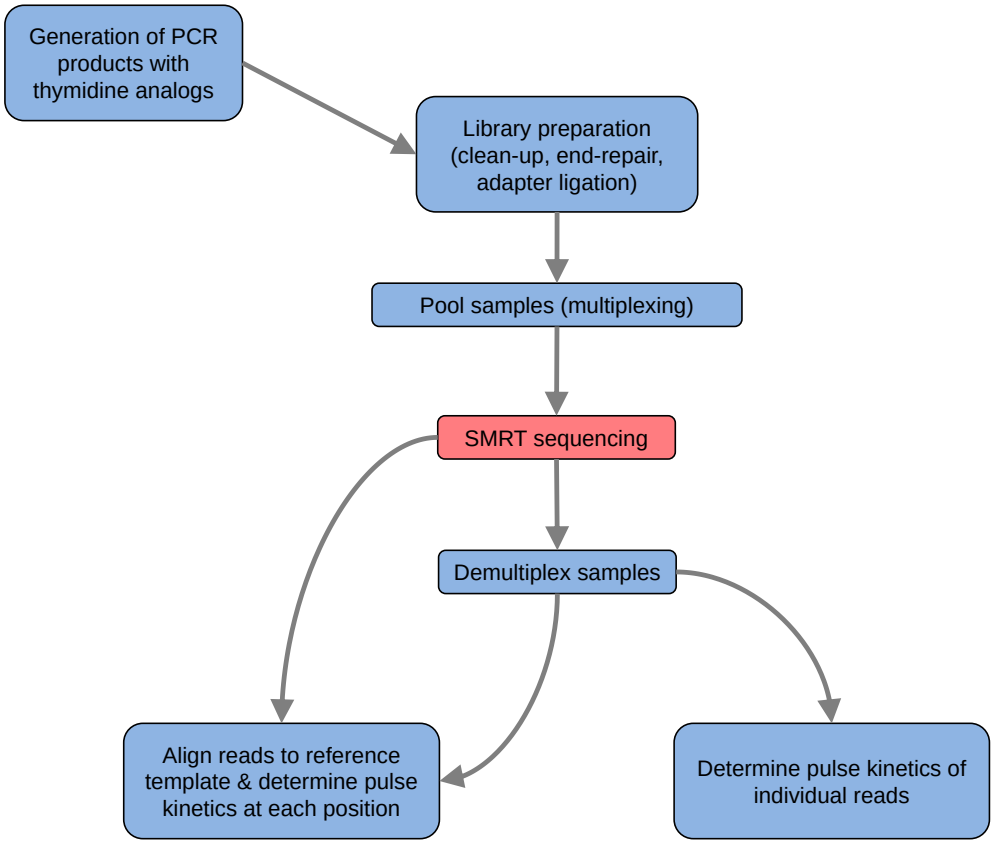


Figure 2.5. Workflow of this study.

Generation of PCR products with thymidine analogs in a 1.6kb DNA region

PCR products containing thymidine analogs in the target sequence were generated (**Figure 2.6**). The region of interest (1.6kb) is comprised of the bases A, C, G, and X (X = T or modified base or both). To facilitate parallel sequencing of multiple samples, barcoded primers were used to keep track of the thymidine analog used. The barcode region of the full length double-stranded PCR product contains only the canonical bases A, C, G, and T. Therefore, the barcode region serves as an internal control.

Alignment of filtered reads to reference template

To determine the polymerase kinetics in the 1.6kb region of interest that contains the thymidine analogs, SMRT-sequenced reads were aligned to the reference template and processed using the PacBio analysis pipeline. Aligned reads were demultiplexed and the IPD ratio for each position in the template was determined (**Figure 2.7**). The IPD ratio (see Materials and Methods for calculation) is generally high in aligned reads from the BrdU, EdU, or IdU libraries (**Table 2.4**). Importantly, libraries that had intermediate amounts of analog (where a 1:1 mixture of analog and T was used in the PCR cocktail) tend to have intermediate IPD ratios. The IPD ratios tend to be higher at positions with the analog (T positions; red) compared to positions without the analog (non-T; blue). The effect of a modified base to influence the IPD at neighboring bases has previously been noted (Flusberg et al., 2010).

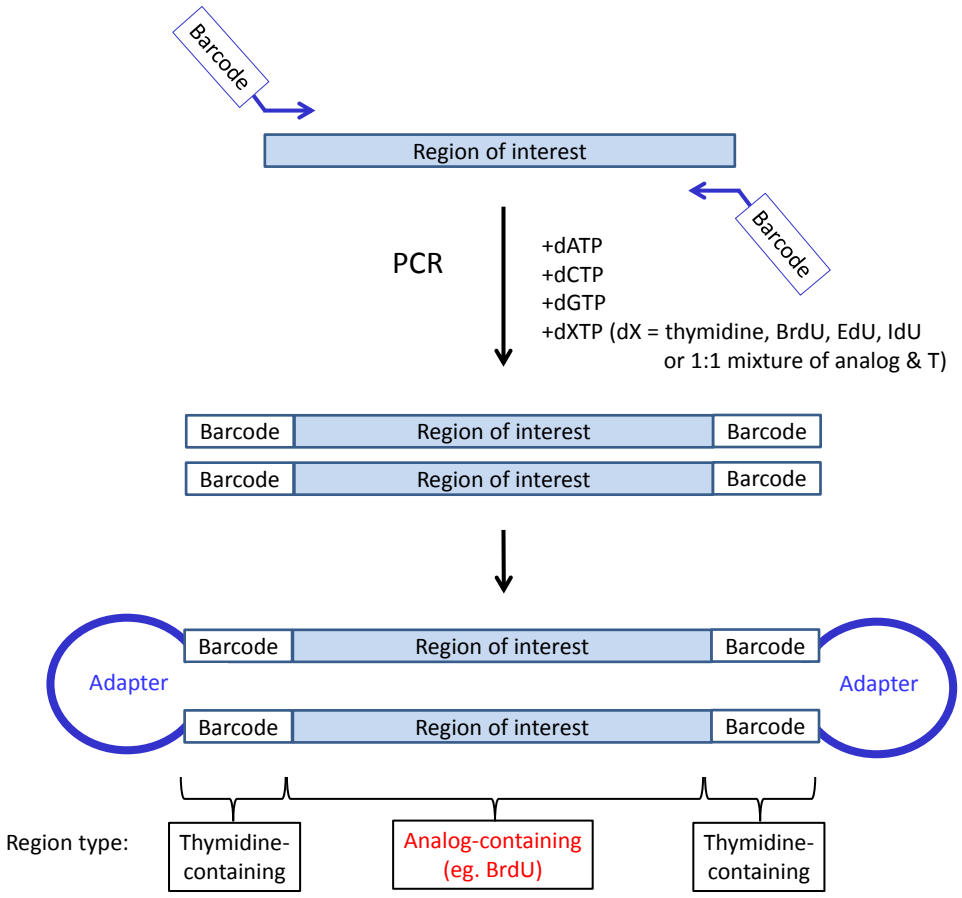


Figure 2.6. Preparation of PCR products that contain thymidine analogs in the 1.6kb region of interest.

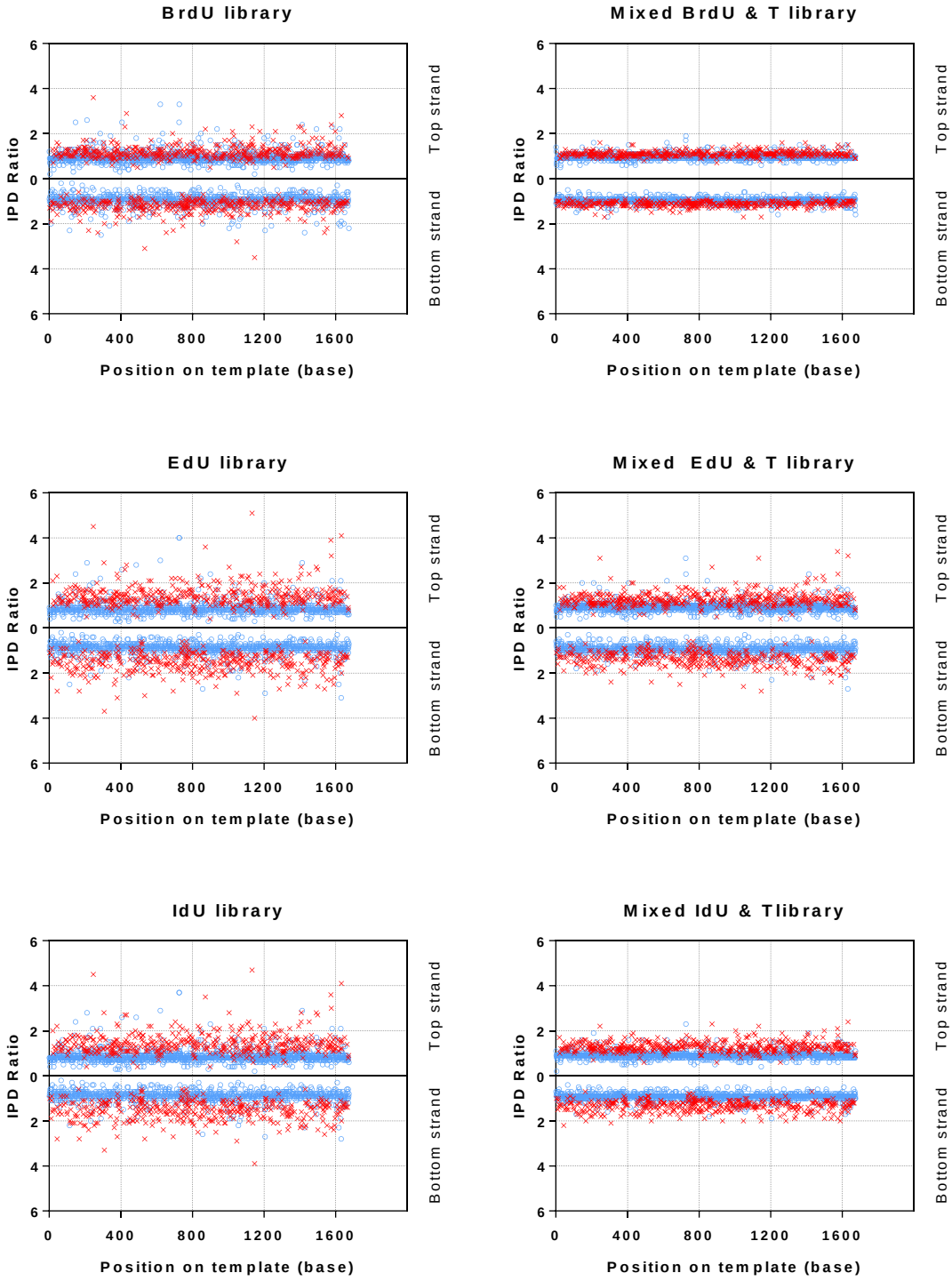
PCR products were used to prepare sequencing libraries, which contain hairpin adapters. In the PCR product, the barcode identifier region contains the canonical bases, A, C, G, and T. The region of interest contains A, C, G, and X, where X = BrdU, EdU, IdU, T, or a 1:1 mixture of analog and T.

Figure 2.7. IPD ratio of thymidine analog-containing libraries.

Raw reads were aligned to the top-strand and bottom-strand of the reference template. The mean IPD at each position was calculated, and the mean IPD was normalized to the control library (T library). The IPD ratio is higher for thymidine or analog positions in the template (red cross), compared to non-thymidine positions (blue circle). The IPD ratios in the pure analog libraries (left column) are generally higher than the IPD ratios in the mixed analog & T libraries (right column).

Figure on next page

Figure 2.7.


BrdU library
Mixed BrdU & T library
EdU library
Mixed EdU & T library
IdU library
Mixed IdU & Tlibrary
IPD Ratio
Top strand
Bottom strand
Position on template (base)
0
400
800
1200
1600
6
4
2
0
2
4
6

Library	Mean of the IPD ratio	
	Thymidine or analog positions (T)	**Non-thymidine positions (non-T)**
BrdU	1.20	0.95
Mixed BrdU & T	1.11	0.97
EdU	1.48	0.87
Mixed EdU & T	1.33	0.93
IdU	1.48	0.88
Mixed IdU & T	1.32	0.91
T (control)	1.00	1.00

Table 2.4. Mean of the IPD ratio is higher in thymidine analog-containing libraries compared to the control library (T library, which contains no analog).

The mean of the IPD ratio in mixed analog & T libraries is intermediate between analog-only libraries and the T control library. The mean of the IPD ratio was calculated as described in Materials and Methods.

A single thymidine-to-EdU substitution increases the IPD

To further validate that thymidine analogs increase the IPD ratio, two synthetic DNA templates were engineered: 1) a control DNA template comprising of the four canonical bases, A, C, G, and T; and 2) an experimental template of the identical sequence, except for the presence of EdU instead of T at position 45. Unique barcodes were included to facilitate sample identification.

As shown in **Figure 2.8**, a single EdU modification increases the IPD ratio to 1.56 at the modified position. The complementary strand at position 45 contains an adenine and its IPD ratio (0.96) indicates no significant change in IPD, as expected. This finding confirms that SMRT-sequencing can detect EdU at a single-base resolution.

Figure 2.8A)

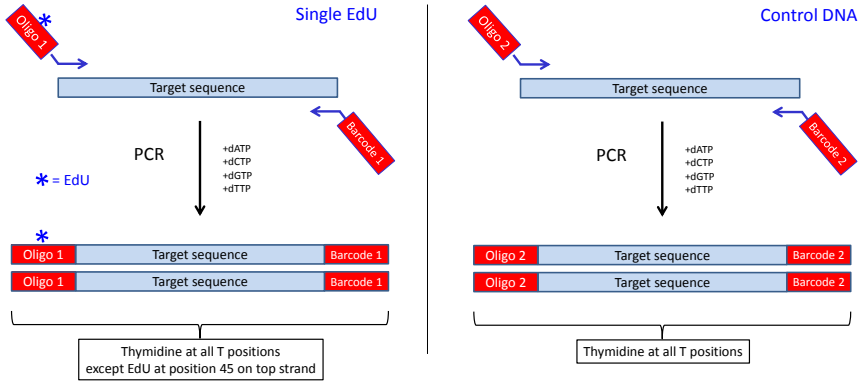


Figure 2.8B)

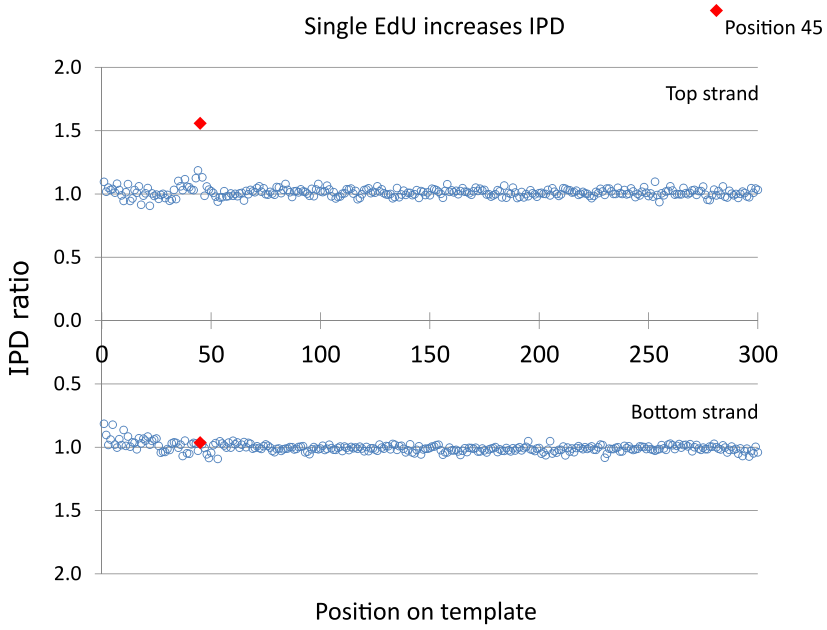


Figure 2.8. Single thymidine-to-EdU substitution increase the IPD ratio.

A) Schematic on preparation of PCR products that contain only a single EdU. To place the EdU in a larger DNA context to facilitate SMRT sequencing, a custom oligonucleotide with EdU is used as a primer in PCR. B) The IPD ratio is high for EdU in the template (Position 45, red diamond on top strand). The complementary strand (bottom strand), which has an A at Position 45, does not have a high IPD ratio. This result indicates that a single thymidine-to-EdU substitution increases the IPD ratio.

Demultiplexing samples with thymidine or analogs in the 1.6kb region of interest

The analysis thus far has examined the average IPD values from a population of aligned reads. To investigate how thymidine analogs affect sequencing kinetics on single DNA molecules, I examined individual reads from libraries with analogs in the 1.6kb target DNA. The samples from different thymidine analog-containing libraries (BrdU, EdU, IdU, and 1:1 mixture of analog & T) were pooled into one SMRT cell for SMRT-sequencing. To examine individual reads, I first demultiplexed the samples to identify reads from each library.

To demultiplex and isolate individual reads for analysis, several demultiplexing strategies were employed. First, the sequencing results were demultiplexed using the PacBio demultiplexing package, *pbbarcode*. This approach assigns barcode pairs to all ZMWs that have high quality polymerase reads (i.e. ZMWs that pass a filter). Regardless of the barcode alignment score, all the ZMWs that pass this filter are assigned a barcode pair. Validation—via BLASTN alignment with the barcodes—failed to confirm the presence of EdU-identifier barcodes in many polymerase reads. Since the read quality is dependent on the IPD, we hypothesized that many EdU reads did not pass the quality filter because of their high IPD values. Consequently, many *bona fide* EdU reads were omitted in the *pbbarcode* demultiplexing algorithm.

In the second demultiplexing approach—using the *ccsBarcode.py* script (John Harting from PacBio, personal communication)—ccsReads were demultiplexed via the Smith-Waterman algorithm. However, this approach recovered many ccsReads that were

assigned to erroneous barcode pairs (e.g. forward BrdU-identifier primer and reverse EdU-identifier primer pair).

To recover more EdU reads, I devised a third approach in which barcode sequences were used to conduct a BLASTN alignment query on all the unfiltered raw reads (that is, on every nucleotide sequence obtained from every ZMW). This BLASTN query yielded 172 unique BLAST hits to the EdU-identifier barcodes, but was significantly less than other analog identifiers (**Table 2.5**). We were intrigued that the EdU samples were an outlier, and we decided to examine individual putative EdU reads in detail.

Identifier barcodes (50 bases) representing libraries containing:	**Non-unique BLASTN hits**
BrdU	43449
EdU	172
IdU	33599
Mixed BrdU & T	44895
Mixed EdU & T	786
Mixed IdU & T	38727
T	49507

Table 2.5. BLASTN hits from the alignment of identifier barcodes to raw reads.

BLASTN was conducted on all base-called sequences in all ZMWs, with the options –word_size 7 -e-value 1e-8. The results show that the non-unique hits are significantly lower for EdU-containing identifiers.

Single molecule analysis shows IPD is low in thymidine regions and high in putative EdU regions

The IPD in a single raw read from the control library—which contains only the canonical bases, A, C, G, and T—is shown in **Figure 2.9**. The IPD value at each nucleotide position of the read is depicted. Throughout this control sample, the IPD is low.

A putative EdU raw read, in contrast, shows a different IPD profile (**Figure 2.10**). Throughout most of this sample, the IPD is high. However, a low IPD valley is present, and its sequence aligns to the barcode and hairpin sequences. The barcode sequence, its complement, and the hairpin sequence contain only the canonical bases A, C, G, and T. In this sample, the hairpin is located near the start of the read (Position 0). However, this is not necessarily the case, because the polymerase reaction may begin before the pulses are recorded (Buschmann et al., 2014). These findings indicate that the IPD is low in thymidine-containing regions and the IPD is high in putative EdU-containing regions.

An unexpected finding in the putative EdU raw read is that the DNA sequence for the high IPD region—as provided from the default basecaller—exceeded 1.6kb, which is the length of the EdU-containing region in the DNA template.

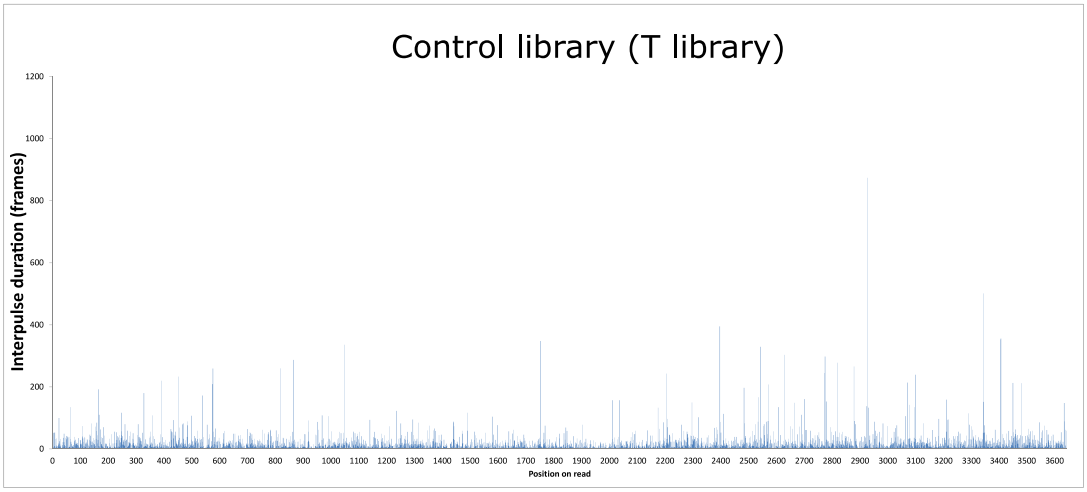


Figure 2.9. Single molecule analysis of a raw read from the control library (T library).

The IPD of this single raw read is relatively low at all template positions.

Figure 2.10. Single molecule analysis of a raw read putatively from the EdU library.

A) The IPD of this raw read is high throughout most of the read. A low IPD valley is found around positions 100-380. B) The IPD of the single read in (A), but showing only the first 500 bases. The reference barcode sequences and the adapter sequence—which do not contain any EdU—map to this low IPD region. C) BLASTN alignment of the barcode sequences and adapter sequences to the read shows a good sequence match.

Figure on next page

Figure 2.10.A)

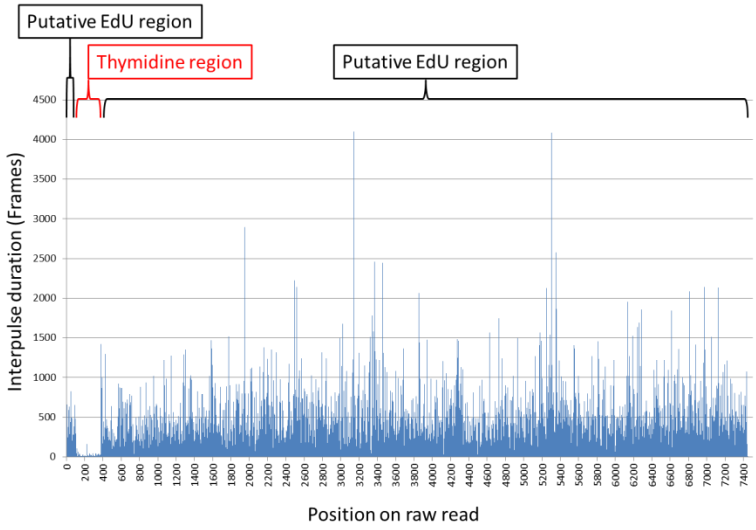


Figure 2.10.B)

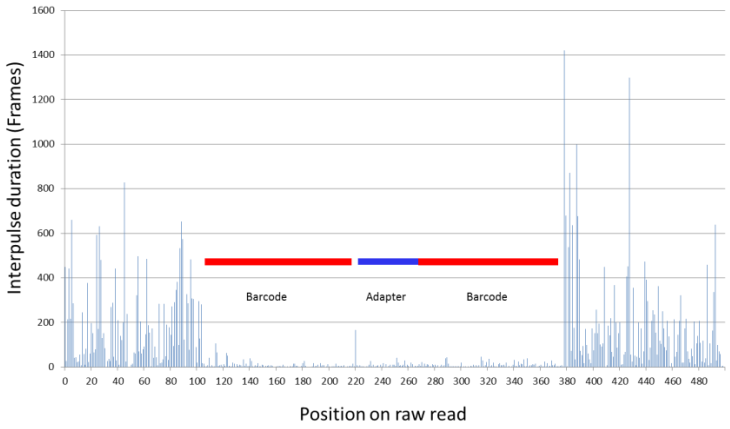


Figure 2.10.C)

```
Barcode   106  TATCGCTTCTCCGG-GCGTGGGCGGTTCTt-gtcgtcctgtgtctgt-gtgccttc      54
               |||||| ||||||| |||||||||||||| ||||||||| |||||| ||||||||
Template  102  TATCGC-TCTCCGGTGCGTGGGCGGTTCTTTGTCGTCCTGGGTCTGTCGTGCCTTC      156

Barcode   53   tggt-gcgtgt-ccgtgtctgt--cgc--tct---gtgctgtgct-cgtctgctctgGTG  4
               ||| |||||| ||||||||||  |||  |||   |||||||||| ||||||||||||||
Template  157  TGATAGCGTGTTCCGTGTCTGTGTCGCACTCTCTTGTGCTGTGCTTCGTCTGCTCTGGTG  216

Adapter   45   TAGAGAGAGT-TG-TTGTTG-TTGCCTCCTCCTCCTTTTCTCTCTCTA             1
               ||||||||||  | |||||| ||||||||||||||||||||||||||||||
Template  221  TAGAGAGAGTCGGTTTGTTGTTTGCCTCCTCCTCCTTTTCTCTCTCTA             268

Barcode   1    TTGGTGgtctc-gtctgctcgt-gtcgtgtctcgctgtc-tgtgcc-t-gtgcgtggtct  55
               ||||| |||| |||||||||| |||||||||||||| |||||| | |||| ||||||
Template  269  TTGGT-GTCTCGGTCTGCTCGTAGTCGTGTCTCGCTGTCGTGTGCCATGGTGCG-GGTCT  326

Barcode   56   t-ccgtgtgtctgtgtcctgctgtTCTTGGCGGGTGCGGGCCT-CTTCGCT          104
               | |||||||||||||||| |||||||||||||||||||||||| |||||||
Template  327  TACCGTGTGTCTGTGTCC-GCTGTTCTTGGCGGGTGCGGGCCTTCTTCGCT          376
```

Effect of EdU on base calling

The finding that only a handful of reads were recovered from the EdU library led to further investigation on individual reads. **Figure 2.11** shows part of the BLASTN alignment of a putative EdU-containing read and the reference template sequence. The alignment suggests insertion errors immediately flanking the EdU position, but not further away. The standard PacBio analysis assigns an IPD for every called base, including the erroneously inserted bases. Thus, although the apparent IPD of the base at the EdU position is 69 frames, a more accurate IPD includes the summation of IPD values from the inserted flanking bases (3321 frames). Since the IPD value is assessed as part of the base caller algorithm, this finding suggests that the apparent scarcity of EdU reads may be due to insertion errors that lower the BLASTN match to the reference sequence. As we used the standard PacBio base-caller without modification, it is plausible that optimization of the base-caller to recognize high IPD values in EdU will provide greater sequencing accuracy.

Figure 2.11A)

```
Query  989   TTTCTTTG--T--GGCCCACATCAATTCCC  1014
              ||| ||||  |  ||||||| ||||| |||
Sbjct  2992  TTTGTTTGACTCCGGCCCACCTCAATCCCC  3021
```

Figure 2.11B)

Reference template (query)	T	G			T			G	G
Base called sequence (subject)	T	G	A*	C*	T	C*	C*	G	G
IPD (frames)	968	1375	45	274	69	2087	846	20	2174
Position on read	2998	2999	3000	3001	3002	3003	3004	3005	3006

* denotes insertion.

Figure 2.11. BLASTN alignment in a single raw read with insertion errors.

This read is putatively from the EdU library, because barcodes specific for the EdU library were identified in this single read. A) Part of the BLASTN alignment of the reference template (Query) to the base-called sequence (Sbjct) from SMRT-sequencing. The alignment shows two inserted bases flanking each side of EdU in the template (T at position 3002). B) The apparent IPD for EdU at position 3002 is 69 frames, but a more accurate value is the summation of the IPD at the neighboring inserted bases (3321 frames).

2.4. Discussion

In this study, I explored the feasibility of using SMRT sequencing to detect thymidine analogs in DNA. I was motivated by the prospect of pulse labeling cells with thymidine analogs to determine the location of replication origins and replication forks on single DNA molecules in a high-throughput manner. Here, I showed that the presence of thymidine analogs can be detected using SMRT sequencing. I prepared synthetic DNA templates where the insert DNA contained either thymidine control, or the thymidine analogs BrdU, EdU, or IdU. Alignment of the reads showed that the mean IPD ratio is higher in templates that contain an analog, compared to thymidine control. Next, I found that even a single thymidine-to-EdU substitution increases the IPD and is detectable via SMRT sequencing. Consistent with EdU increasing the IPD, single molecule analyses of individual reads suggest putative EdU-containing regions have high IPD, while EdU-devoid regions have low IPD. These findings are consistent with the polymerase progressing with prolonged timing at EdU regions, but can return to baseline kinetics in thymidine-containing (EdU-absent) regions.

For my single molecule analysis of individual reads, the PCR products I prepared had a 1.6kb EdU-containing region of interest. The region of interest is flanked by barcode regions that do not have EdU. Under ideal conditions, SMRT sequencing would yield a high IPD region of exactly 1.6kb, separated by low IPD regions that characterize the barcode. In the handful of reads studied, the high IPD region extends beyond the 1.6kb range. Moreover, I was unable to align with high accuracy the base-called sequence

to the control region of interest sequence. This may be attributed to the base-calling algorithm, which has not been optimized for EdU detection. Consequently, the high IPD at EdU-containing regions may have led to insertion errors (**Figure 2.11**) that made it difficult to obtain a BLASTN alignment to the control amplicon.

Even if the base-caller is not optimized for calling analog-containing regions with single base resolution accuracy, SMRT sequencing can still be applied to map replication origins and replication forks. For instance, an investigator could use an *E. coli* strain (Marsh and Worcel, 1977) or *S. cerevisiae* strain (Vernis et al., 2003) that have mutations in the thymidylate synthase gene, which requires exogenous thymidine supplementation for growth. The strains can be synchronized in G1 phase and released into a thymidine analog-containing medium in the presence of hydroxyurea. Hydroxyurea inhibits ribonucleotide reductase, thus limiting the endogenous dNTP pool (Elford, 1968; Timson, 1975). Cells in G1 phase released into S-phase with hydroxyurea can initiate DNA replication at early origins, and replication forks can progress for a short duration before the forks are arrested (Santocanale and Diffley, 1998). After hydroxyurea arrest, cells can be washed out and re-suspended in thymidine-rich medium to allow for replication completion. This process results in replicated double-stranded DNA, with the mother strand containing only the four canonical bases, A, C, G, and T, and the daughter strand containing small pulses of the thymidine analog at early origins. Genomic DNA can be processed, SMRT sequenced, and reads with high IPD ratios can be selected for analysis. Ligating hairpins to the DNA yields templates that contain a mother strand and a daughter strand. Since the mother strand is comprised of the four canonical bases, its

sequence can be used to map each read to the reference genome. Thus, even if the base-caller cannot accurately provide the sequence of the analog-containing regions in the daughter strand, the unmodified bases in the mother strand can be used for mapping.

The standard PacBio RS II analysis pipeline takes raw fluorescence traces and converts them to a base-called sequence for each read. In the standard workflow, the end-user is provided with the base-called sequences, and not the fluorescence traces. To learn more about polymerase sequencing kinetics with thymidine analogs, it would be helpful to examine the raw fluorescence traces. Such information has not been made available at the time of writing, but such data could facilitate in optimizing the base-calling algorithm to provide higher accuracy in base-calling regions that contain thymidine analogs.

In addition to identifying replication origins on single DNA molecules, being able to map the location, velocity, and direction of replication forks in a high-throughput manner would advance our understanding of replication kinetics. Detection of sequential pulses of two independent analogs can provide fork directionality. Both 5-methylcytosine and N6-methyladenine—analogs not derived from thymidine—have been detected via SMRT sequencing and have high IPD values (Flusberg et al., 2010). My study can be extended to examine the kinetics of incorporating other non-thymidine analogs, which may be used in conjunction with EdU. Optimization of the base calling algorithm may improve the base calling accuracy to allow the discrimination of two types of modified bases. This would allow the identification of replication fork directionality on single DNA molecules. Moreover, it may provide further insight into whether replication fork

velocities are uniform (Sekedat et al., 2010) or not (Conti et al., 2007). SMRT sequencing of pulse labeled DNA can thus advance the field of replication kinetics.

Chapter III: Optical mapping of DNA replication origins in nanofluidic channels

3.1. Introduction

DNA replication initiates at sites termed origins (Huberman and Riggs, 1968), but their location and behavior in higher eukaryotes are not well understood. Existing single molecule approaches to study replication origins and replication kinetics are low throughput. Here, we explored the feasibility of using the BioNano Genomics Irys platform as a high throughput technology to map replication origins and replication forks on single DNA molecules.

The BioNano Genomics Irys platform is a next-generation mapping technology capable of determining the genomic source of individual DNA molecules. The technology is compatible with an optical approach that includes three key processes: 1) labeling DNA molecules with fluorescent nucleotides at specific sequences (nick sites); 2) detection of linear DNA molecules in nanofluidic channels; and 3) alignment of DNA molecules with fluorescent labels to a reference genome.

To prepare DNA for optical mapping, fluorescent labels are introduced into double-stranded DNA at specific sequences via nick translation (Lam et al., 2012; Luzzietti et al., 2012; Xiao et al., 2007). In nick translation, a sequence-specific nicking endonuclease is used to produce an incision on one strand of double-stranded DNA. Nicking endonucleases function similar to restriction enzymes, but nicking endonucleases produce single-stranded nicks whereas restriction enzymes produce double-stranded nicks. Consequently, the use of nicking endonucleases creates a higher frequency of intact double-stranded DNA molecules with multiple sequence-specific nicks. After the

nicking reaction, a DNA polymerase is used to incorporate fluorescent nucleotides at the nick site. This process results in double-stranded DNA with fluorescent labels at the nick sites. Fluorescently labeled DNA can then be imaged in nanofluidic systems (Das et al., 2010; Jo et al., 2007).

A centerpiece of the Irys platform is a nanofluidic chip (IrysChip) that contains thousands of parallel channels, each 45nm in diameter (Lam et al., 2012). Each nanochannel is designed to allow single DNA molecules to flow through, but narrow enough such that DNA molecules retain a linear form. For optical mapping, DNA molecules are loaded onto the nanofluidic chip and subjected to an applied electric field. Electrophoresed DNA molecules travel through progressively narrower conduits interspersed with nanoscale obstacles that uncoil the DNA. This funneling process allows DNA molecules to enter the spatially confined nanochannels and retain a linear form. Hundreds of parallel DNA molecules flowing in nanochannels can be automatically captured in each camera frame-of-view, thus facilitating the rapid collection of data. Multicolor fluorescence detection allows the visualization of the DNA molecules with the fluorescent tags. Imaging rates of up to 5Gb of DNA per hour have been reported, making this a high throughput technology (Hastie et al., 2013).

Since the nicking endonuclease is sequence-specific, the spacing between the nick labels—for DNA molecules that are linear and uniformly stretched—is sequence-dependent. The nick labels on linear DNA molecules thus form a barcode pattern, which can be aligned to the known pattern of a reference genome (**Figure 3.1a-c**). This process enables DNA molecules to be mapped to a reference genome.

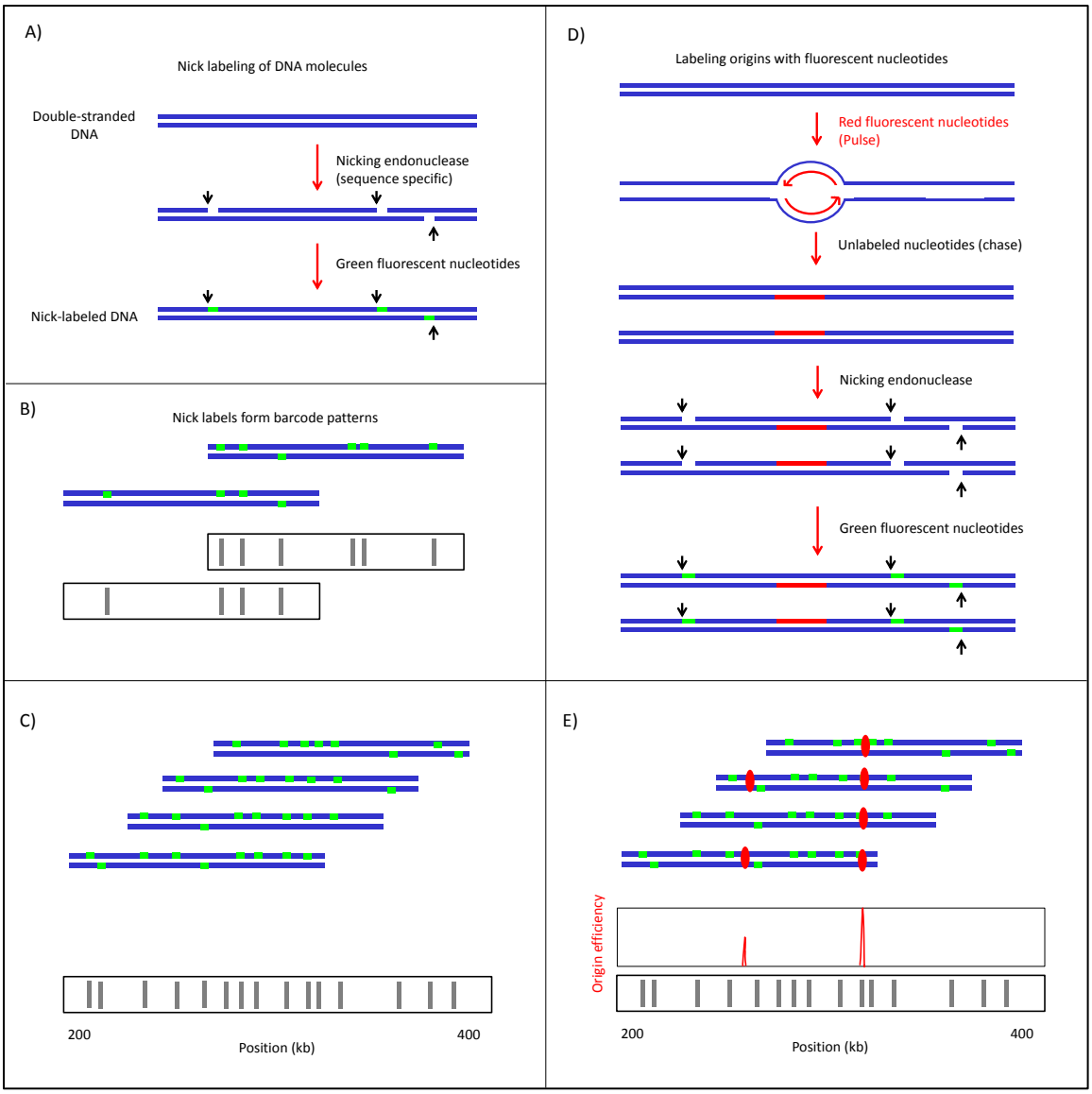


Figure 3.1. Nick labeling of DNA for optical mapping.

A) Double-stranded DNA is treated with a nicking endonuclease, which recognizes a specific sequence and introduces an incision on one strand of the DNA. After a nick is introduced, a DNA polymerase is used to label the nick site. **B)** The nick labels form barcode patterns. **C)** The nick labels can be used to align to DNA molecules to the reference genome. **D)** Direct labeling of replicating origins. Red fluorescent nucleotide pulses are used to label the sites of active replication. To produce double-stranded DNA, unlabeled nucleotides are used to chase. Next, nick labels are introduced to map the DNA molecules. **E)** Alignment of DNA molecules with replication origins to the reference genome. The green labels facilitate mapping of DNA molecules. The red labels mark the site of replication origins.

The Irys platform has been used to study genome structural variations and re-arrangements (Cao et al., 2014; Mak et al., 2016). However, using this technology to map DNA replication origins remains a novel endeavor. The ability to optically map replication origins and replication forks in a high throughput manner would revolutionize the field of replication kinetics. It would allow the rapid study of how replication origins, replication fork densities, and replication fork speeds are affected upon genetic mutations, varying environmental conditions, and the presence of small chemical molecules and mutagens.

An illustration of how replication origins might be mapped using the Irys platform is shown in **Figure 3.1d-e**. In this direct labeling of replication origins approach, live cells are arrested so that no replication is occurring. Next, cells are allowed to begin replication in the presence of a small dose of red fluorescent nucleotides. The red fluorescent nucleotides thus mark the location of the origins. Cells are washed out to remove unincorporated red fluorescent nucleotides, and allowed to complete replication with unlabeled nucleotides. Afterwards, double-stranded DNA is nick translated with green fluorescent nucleotides and loaded on a nanofluidic chip for imaging and mapping. Molecules with red fluorescence are identified as origin-containing molecules, and their locations are determined via alignment.

Direct labeling of replication origins with fluorescent nucleotides, however, remains a challenge *in vivo*. Fluorescent nucleotides are relatively bulky and charged, making them difficult to be transported into live cells. Transient transfection of fluorescent nucleotides can be used to mark ongoing replication tracks. Cell sorting via

FACS can select for transfected eukaryotic cells in S-phase (replicating cells) (Gilbert, 2010). However, this provides information on ongoing replication tracks and not the replication start site. To map replication origins, cell cycle synchronization is generally required to obtain a homogenous population of cells at the start of S-phase. However, cell cycle synchronization adds complexity to a transient transfection approach, because the recovery time after transfection would need to be coordinated to ensure that fluorescent nucleotides are incorporated in only one specific cell cycle.

An alternative pulse labeling approach is to use smaller cell-permeable nucleoside analogs, isolate genomic DNA, and then detect the analogs using fluorophores. 5-Bromo-2'-deoxyuridine (BrdU) is a cell-permeable thymidine analog used to study replication kinetics. BrdU can be detected by immunofluorescence, but DNA denaturation is classically required for antibody access (Gratzner, 1982). The need for DNA denaturation and the large size of antibodies relative to nanochannels makes BrdU detection incompatible with the Irys platform. Similarly, antibody detection of the thymidine analogs 5-chloro-2'-deoxyuridine (CldU) and 5-iodo-2'-deoxyuridine (IdU) share the same limitations as BrdU.

In recent years, 5-ethynyl-2'-deoxyuridine (EdU) has emerged as an alternative to BrdU for DNA labeling. EdU contains a terminal alkyne that can be detected using click chemistry. The term "click chemistry" was coined to describe highly specific and efficient chemistry reactions that can be used to detect biomolecules (Kolb et al., 2001). The prototypic click chemistry reaction is the copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) reaction (Rostovtsev et al., 2002). In CuAAC, a terminal alkyne

reacts with an azide group in the presence of copper(I) catalyst to form a 5-membered ring. EdU, which contains a terminal alkyne, can be conjugated to an azide-containing fluorophore in a CuAAC reaction (Salic and Mitchison, 2008). Thus, EdU can be detected using CuAAC click chemistry. EdU detection in whole cells does not require DNA denaturation or the use of antibodies, making it a candidate for labeling origins. The CuAAC reaction requires a copper(I) catalyst, which is commonly derived from the reduction of copper(II) in aqueous $CuSO_4$ by ascorbic acid (Meldal and Tornøe, 2008). While copper(II) itself is not toxic at experimental concentrations, its reduced form—copper(I)—can lead to cytotoxicity, raising the question of whether EdU detection is compatible with optical mapping (Aruoma et al., 1991; Buettner and Jurkiewicz, 1996; Cappella et al., 2013; Cross et al., 2003; Imlay and Linn, 1988; Liu et al., 2006; Urbañski and Beresewicz, 2000).

A variant of the classic CuAAC click chemistry reaction is the strain-promoted azide-alkyne cycloaddition (SPAAC) reaction (Agard et al., 2004; Baskin et al., 2007). SPAAC was developed to overcome the cytotoxicity in CuAAC. In SPAAC, a ring-strain is used to promote the reaction between an alkyne and an azide. Structurally, the alkyne is destabilized as a member of a non-aromatic ring (e.g. cyclooctyne). The use of electron-withdrawing groups (such as fluorine in difluorocyclooctyne) can further destabilize the alkyne (Baskin et al., 2007). Other commonly used cyclooctyne derivates are the dibenzocyclooctynes (DIBO or DBCO) (Ning et al., 2008). Fluorophores attached to DIBO or DBCO are commercially available, including DIBO-Alexa Fluor 594 and

DBCO-PEG4-TAMRA. SPAAC does not require a copper(I) catalyst, and thus addresses the concern of copper(I) cytoxicity.

SPAAC reactions require two substrates: a ring-strained alkyne and an azide. The development of 5-azidomethyl-2'deoxyuridine (AmdU) has facilitated DNA label detection via SPAAC. AmdU is a cell-permeable thymidine analog with a reactive azide group (Neef and Luedtke, 2014). AmdU-labeled DNA can be detected in two ways: 1) with an alkyne-containing fluorophore in the presence of copper (CuAAC); or 2) with a ring-strained alkyne-containing fluorophore in the absence of copper (SPAAC).

Detection of replication origins on DNA molecules

To detect DNA replication origins on single DNA molecules that can be imaged in nanofluidic channels for genome mapping, we devised three approaches.

The first approach involves mapping replication bubbles ("bubble mapping"). Here, DNA molecules are stained with a DNA binding fluorescent dye (e.g. YOYO-1, a fluorescent DNA intercalator). A replication bubble has twice the DNA content compared to unreplicated DNA. Therefore, the fluorescence intensity at a replication bubble should be 2x more than its neighboring unreplicated region. Thus, I hypothesized that by mapping DNA molecules for regions of double fluorescence intensity, we can identify replication origins (**Figure 3.2**). Consistent with this hypothesis, DNA molecules stretched on glass coverslips—in a process called molecular combing or DNA combing—were found to have twice the YOYO-1 fluorescence intensity at the replication tracks (De Carli et al., 2016). A key advantage of this approach is that it requires no pulse labeling of nucleoside analogs, which may have an inhibitory and/or toxic effect on proteins involved in endogenous nucleoside and nucleotide metabolism and incorporation.

Figure 3.2A)

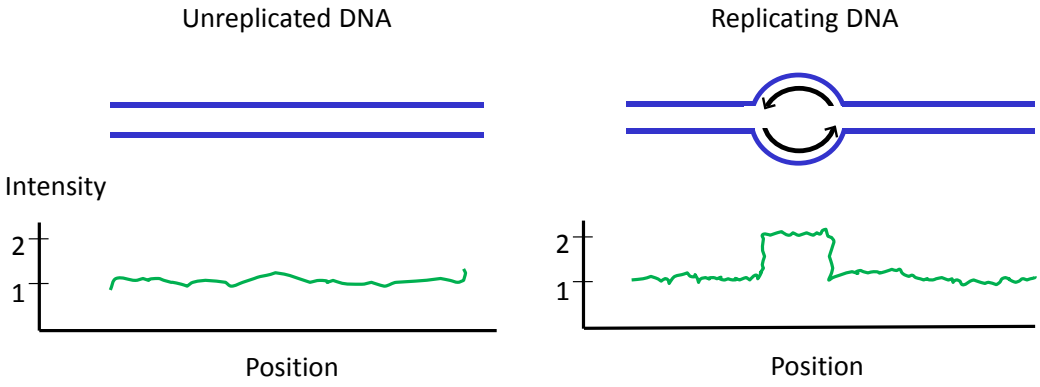


Figure 3.2B)

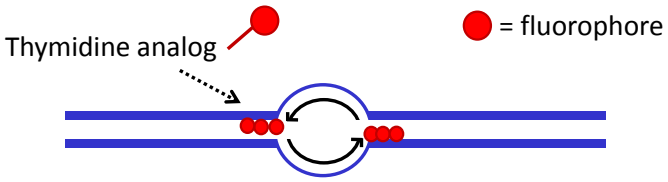


Figure 3.2. Detection of replication origins on single DNA molecules.

A) Replication bubble detection based on DNA fluorescence intensity. Replication bubbles have twice the copy number of DNA and thus are expected to have twice the YOYO-1 DNA fluorescence intensity. **B)** Labeling replication forks with fluorescent nucleotides.

The second approach (“bubble labeling”) involves labeling replication forks with fluorescent nucleotides. Replication forks are located at the ends of the replication bubble. Here, the leading strand has a free 3’-hydroxyl substrate that may be extended by a DNA polymerase. This reaction is akin to primer extension. Replication forks can be labeled using fluorescent nucleotides. Genomic DNA—either in solution or embedded in agarose plugs—can be used provided that replication forks are intact.

The third approach (“nucleoside pulse labeling”) involves pulse labeling live cells with cell-permeable nucleoside analogs that can then be conjugated to fluorescent dyes via click chemistry (CuAAC or SPAAC).

In this study, all three approaches were investigated. As proof-of-concept, *Escherichia coli*, which has only one origin of replication, was used to determine the effectiveness of optically mapping origins (Bonhoeffer and Gierer, 1963). The *E. coli* origin is well-studied, well-defined, and site-specific, located at the 3.92Mb position (Blattner et al., 1997; Marsh, 1978; Marsh and Worcel, 1977; Yasuda and Hirota, 1977). Our results suggest that all three approaches are plausible strategies to combine with optical mapping of DNA molecules. The approaches can ultimately be extended to study replication origins and replication kinetics on single DNA molecules from both prokaryotes and eukaryotes.

3.2. Materials and Methods

Preparation of *E. coli* DNA (bubble-enriched samples and unreplicated control)

The *Escherichia coli* PC5 strain (F-, *leuB6*(Am), *λ*-, *thyA47*, *rpsL153*(strR), *dnaA5*(ts), *deoC3*) was obtained from the *E. coli* Genetic Stock Center (CGSC Strain #5944) (Wechsler and Gross, 1971).

E. coli samples were prepared for flow cytometry and optical mapping as outlined in **Figure 3.3**, using a temperature-sensitive arrest and release protocol previously described (Marsh and Worcel, 1977).

Flow cytometry to monitor DNA replication

EdU incorporation into DNA was detected via CuAAC click chemistry using Click–iT EdU Alexa Fluor 488 Imaging Kit (Invitrogen). Reactions were performed at room temperature. Cells fixed in 90% methanol were washed twice in PBS, permeabilized in 0.5 % Triton X-100 in PBS (20 minutes), washed twice in PBS, and treated with the reaction cocktail (30 minutes) per manufacturer instructions. The Click-iT reaction cocktail (500uL) was prepared with the following reagents: 1x Click-iT reaction buffer (430uL); $CuSO_4$ (20uL), Alexa Fluor 488 azide (1.2uL); Reaction buffer additive (50uL).

Cells were then washed four times in PBS and re-suspended in sheath fluid. Flow cytometry detection of the Alexa Fluor 488 signal was performed on a FACScan (Beckton-Dickinson Immunocytometry, San Jose, CA, USA).

Optical mapping of DNA molecules

Glycerol stocks of *E. coli* samples (~3 x 10^9 cells per tube) were sent to BioNano Genomics (San Diego, CA) for processing and optical mapping on the Irys platform. Site-specific nicks introduced with *NtBspQI* (specific for 5'-GCTCTTC-3') were labeled with green fluorescent dUTPs using *Taq* DNA polymerase (72^{o}C for 1 hour). DNA was stained with YOYO-1.

Labeling of replication forks in solution prior to nick labeling

E. coli genomic DNA was treated +/- DNA ligase to repair non-specific pre-existing nicks. Red fluorescent dUTPs were incorporated using *Taq* DNA polymerase (72^{o}C for 1 hour). Drop dialysis was performed to remove unincorporated free red fluorescent dUTPs. Site-specific nicks introduced with *NtBspQI* were labeled with green fluorescent dUTPs using *Taq* DNA polymerase (72^{o}C for 1 hour). DNA was stained with YOYO-1.

Rolling circle amplification

Rolling circle amplification assay was performed with M13mp18 circular single-stranded DNA (New England Biolabs) as the template. To increase the reaction yield, two forward primers (VL-086 and VL-088) targeting diametrically opposite positions on the circular template were used. Mid-way through the extension reaction, a reverse primer (VL-087) was added to make the products double-stranded. An optional

restriction digestion step with EcoRI and PacI was used to fragment the end product, but the PacI digest was found to have low activity in this reaction mixture.

EdU and AmdU labeling in mammalian cells

Mouse erythroleukemia cells with an integration of the herpes simplex virus thymidylate synthase gene (MEL-HSV-tk cells) were obtained from Dr. Mirit Aladjem (Feng et al., 2005).

HeLa cells and MEL-HSV-tk cells were cultured in Dubecco's Minimum Essential Media (DMEM) supplemented with 10% fetal bovine serum (FBS), 1% penicillin and streptomycin under standard culture conditions (37^oC, 95% air, 5% CO_2). Growing cells were treated with 25uM EdU or 25uM AmdU or DMSO control overnight, washed with DPBS, and harvested.

EdU incorporation detection via CuAAC was done using Click –iT EdU Alexa Fluor 594 Imaging Kit (Invitrogen). AmdU was synthesized at UMMS (Ryan Holmes) or purchased from Sigma. AmdU incorporation detection via CuAAC was done with Alexa Fluor 594 Alkyne (the other reaction components were obtained from the Click–iT Imaging Kit (Invitrogen)). AmdU incorporation detection via SPAAC was done using Click-iT Alexa Fluor 594 DIBO Alkyne (Invitrogen) or Dibenzylcyclooctyne-PEG4-5/6-Tetramethylrhodamine (DBCO-PEG4-TAMRA; JenaBiosciences) according to the manufacturer’s protocol for the Click-iT Alexa Fluor 594 DIBO Alkyne kit.

Subsequent steps were performed at room temperature: harvested cells were fixed in 3.7% formaldehyde in PBS (15 minutes), washed twice in DPBS, permeabilized with

0.5% Triton X-100 in DPBS (20 minutes), washed twice with DPBS, and treated with the reaction cocktail (30 minutes) per manufacturer instructions. Cells were washed four times in DPBS and nuclear-stained with Hoechst 33384 at 1ug/mL (ThermoFisher Scientific).

Molecular combing of DNA fibers (DNA combing)

HeLa cells treated with 25uM EdU or DMSO control overnight were washed, trypsinized, harvested, and processed for fluorophore conjugation via copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) according to the manufacturer's protocol for Click –iT EdU Alexa Fluor 594 Imaging Kit (Invitrogen). After CuAAC, cells in DPBS (0.1 million to 1 million cells per plug) were embedded in 1.5% low melting agarose plugs (Bio-Rad) to yield a final 0.75% agarose concentration. Agarose plugs were then digested in Proteinase K solution (2mg/mL Proteinase K, 1% sarkosyl, 10mM Tris pH 7.5, 50mM EDTA) in a 50^{o}C waterbath for 12 hours, after which four additional 12-hour incubations with fresh Proteinase K solution each time was performed. Next, agarose plugs were washed four times in TE solution (10mM Tris pH 7.5 and 50mM EDTA in water). Agarose plugs were stored at 4^{o}C in TE solution.

To prepare silanized glass coverslips for molecular combing, glass coverslips were processed in a plasma cleaner (Harrick Plasma, Ithaca, NY) and coated with octenyltrichlorosilane (Sigma Aldrich) overnight in a sealed glass chamber (Iyer et al., 2017).

For DNA combing, each agarose plug was placed in a 2mL round bottom microcentrifuge tube containing 500uL TE, 0.2uL YOYO-1, 480uL 0.5M MES pH 5.4, 420uL H2O. The plugs in solution were incubated at 65^{o}C for 30minutes, shifted to 42^{o}C for 30minutes, and digested with 40uL beta-agarase mix (4uL beta-agarase, 4uL 10x NEB beta-agarase buffer, 32uL H2O; New England Biolabs) at 42^{o}C overnight. The resulting solution contains DNA with melted and digested agarose. Each tube was microcentrifuged at 800g for 5minutes at room temperature and the top 1mL to 1.4mL of the solution was transferred using cut pipette tips to a DNA combing solution vessel. A DNA combing instrument (Genomic Vision, France) was used to dip silanized coverslips into the vessel for 5minutes, after which the coverslips were withdrawn at a constant speed to facilitate the stretching of DNA molecules. Coverslips with DNA fibers were visualized for YOYO-1 signal using a Zeiss Axioskop 2Plus microscope at 100x magnification.

Pulsed field gel electrophoresis

DNA in low melting agarose plugs (Bio-Rad) were placed in 1% TBE agarose gels. Pulsed field gel electrophoresis was performed in a CHEF-DR II instrument (Bio-Rad) with the following conditions: 0.5x TBE buffer, 16 hour runtime, 6 V/cm voltage gradient, 5-60 seconds pulse parameters, 15^{o}C buffer temperature. After electrophoresis, the agarose gel was stained in 0.5ug/mL ethidium bromide in 0.5x TBE buffer, de-stained in water, and imaged.

Effect of BTTP ligand on copper(I)-mediated DNA damage

Serial dilutions of $CuSO_4$ and the ligand BTTP (3-(4-((bis((1-*tert*-butyl-1*H*-1,2,3-triazol-4-yl)methyl)amino)methyl)-1*H*-1,2,3-triazol-1-yl)propanol) were prepared to obtain $CuSO_4$-BTTP mixtures with final $CuSO_4$ concentrations of 1mM, 0.37mM, 0.1mM, 0.037mM, 0.01mM, 0.0037mM, and 0.001mM. BTTP from Dr. Peng Wu (AECOM) was used in a 4:1 ratio for [BTTP]:[$CuSO_4$]. For each $CuSO_4$-BTTP dilution, 1ug of Lambda DNA (New England Biolabs) was treated with Click-iT reaction cocktail mixes (Invitrogen): 1x Reaction Buffer (21.4uL), Reaction buffer addictive (2.5uL), and $CuSO_4$-BTTP mixture (1uL). Reactions were performed at room temperature for 30minutes, and aliquots were loaded for DNA gel electrophoresis in a 1% TAE agarose gel. The following concentrations of $CuSO_4$ were used: Lanes 1 and 2 (1mM); Lanes 3 and 4 (0.37mM); Lanes 5 and 6 (0.1mM); Lanes 7 and 8 (0.037mM); Lanes 9 and 10 (0.01mM), Lanes 11 and 12 (0.0037mM); Lanes 13 and 14 (0.001mM); Lanes 15 and 16 (0mM).

Effect of BTTP ligand on CuAAC-mediated whole cell staining

HeLa cells were treated with 25uM EdU or DMSO overnight, harvested, fixed in 3.7% formaldehyde, permeabilized in 0.5% Triton X-100, and treated with Azide-Alexa Fluor 594 to conjugate the EdU via the CuAAC reaction. This reaction was performed according to the manufacturer's protocol in the Click–iT EdU Alexa Fluor 594 Imaging Kit (Invitrogen). To determine the effect of BTTP on CuAAC efficiency, a $CuSO_4$-BTTP premix was used containing 2mM $CuSO_4$ and 8mM BTTP. BTTP addition dampened the fluorescence intensity. It should be noted that although Cu^{2+} in $CuSO_4$ is catalytically inactive for azide-alkyne cycloaddition reactions, the presence of ascorbic acid in the cocktail reaction mixture reduces Cu^{2+} to the catalytically active Cu^{1+} form.

3.3. Results

The results are presented according to the three approaches used to investigate the feasibility of using the BioNano Genomics Irys platform to map replication origins and replication forks.

3.3.1. Strategy #1: Bubble mapping (DNA fluorescence intensity approach)

3.3.2. Strategy #2: Bubble labeling (replication fork labeling)

3.3.3. Strategy #3: Nucleoside pulse labeling

3.3.1. Strategy #1: Bubble mapping (DNA fluorescence intensity approach)

The first approach is to identify origins in nanofluidic channels via DNA fluorescence intensity. Compared to unreplicated regions, replication bubbles (origins) have twice the DNA content and thus should have twice the DNA fluorescence intensity.

Preparation of *E. coli* DNA for fluorescence intensity detection

To map the *E. coli* origin of replication using single DNA molecules in nanofluidic channels, two sets of *E. coli* genomic DNA were prepared for staining: replication bubble-enriched DNA and unreplicated control DNA. The *E. coli* PC5 strain has two relevant mutations: 1) a mutation in *thyA*, the thymidylate synthase gene, that makes the strain require exogenous thymidine for growth; and 2) a temperature sensitive mutation in *dnaA*, the bacterial replication initiator gene (Marsh and Worcel, 1977; Wechsler and Gross, 1971). At the restrictive temperature (40^{o}C), the PC5 strain cannot initiate replication. However, any replication events initiated prior to the arrest can continue to completion (Marsh and Worcel, 1977).

Classification of YOYO-1 stained DNA molecules

DNA samples were prepared from *E. coli* cultures in (**Figure 3.3a**). Flow cytometry analyses were consistent with cell cycle arrest (**Figure 3.3b**). DNA molecules were stained with YOYO-1 (a DNA-binding fluorescent dye), nicked with *NtBspqI*, and the nicks were labeled with green fluorescent nucleotides. DNA molecules loaded onto a nanofluidic chip were imaged using the BioNano Genomics Irys platform. YOYO-1 fluorescence intensities along each single DNA molecule were quantitated.

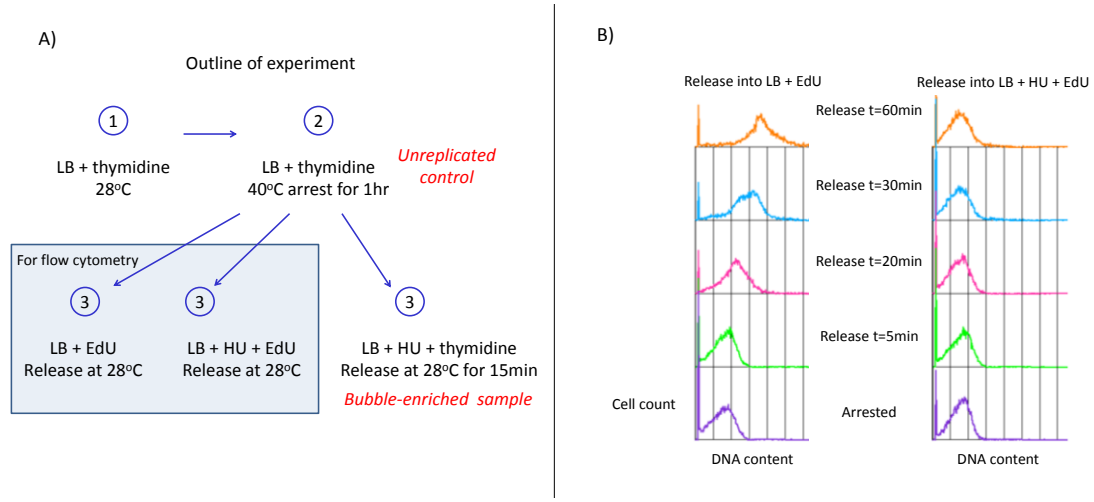


Figure 3.3. Preparation of *E. coli* samples with replication bubbles for flow cytometry and for optical mapping.

A) Outline of experiment. The thymidine-requiring *E. coli* PC5 *thyA47 dnaA5*(ts) strain was arrested at the 40°C (restrictive temperature) for 1 hour in thymidine-rich medium. For flow cytometry analysis, after washout to remove thymidine, cells were released into medium containing either EdU (left) or EdU with hydroxyurea (right). Samples at timepoints were cooled on ice and stored in methanol for flow cytometry analysis. For optical mapping of bubble-enriched samples, arrested cells were released into thymidine with hydroxyurea at t=15mins. For optical mapping of unreplicated control, cells arrested at the restrictive temperature for 1 hour were used. Samples for optical mapping were treated with sodium azide and flash frozen with 30% glycerol in liquid nitrogen. LB, lysogeny broth; EdU, 5-ethynyl-2'-deoxyuridine (100uM); HU, hydroxyurea (1M). **B)** Flow cytometry analysis of EdU content detected with azide-Alexa Fluor 488 is consistent for the expected cell cycle arrest and release.

DNA molecules that are >100kb in length were classified as having feature(s) or no features. A feature is defined as a region of at least double YOYO-1 fluorescence intensity relative to the DNA backbone. In the bubble-enriched sample, 12% of DNA molecules that are >100kb in length had feature(s). In the unreplicated control, 21% of DNA molecules that are >100kb in length had feature(s) (**Table 3.1**). DNA topological events such as DNA knots and DNA folds, which show regions of increased YOYO-1 fluoresence intensity, have been reported to contribute to the noise in the Irys platform (Reifenberger et al., 2015).

	Molecule classification (> 100kb molecules)	
	No features	Feature(s)
Unreplicated control (n=13033)	79%	21%
Bubble-enriched sample (n=13024)	88%	12%

Table 3.1. Classification of DNA molecules >100kb.

Unreplicated control and bubble-enriched *E. coli* samples were nick-labeled and stained with YOYO-1 DNA-binding fluorescent dye. DNA molecules in nanochannels were imaged and classified as either having features or no features. Features are defined as regions of double YOYO-1 intensity relative to the DNA backbone.

Distinct peak in coverage at the *E. coli* origin of replication

DNA molecules mapped to the *E. coli* genome were unevenly distributed throughout the genome. This could be due to genomic coldspots where the barcode pattern of nick sites makes alignments difficult. For example, at places where the nick sites are sparse, aligning DNA molecules can be difficult. Alternatively, the genomic coldspots may be attributable to a high density of nick sites that can lead to DNA molecule scission and result in short DNA molecules that are not mapped. To account for coverage bias, we took the coverage of DNA molecules with feature(s), and subtracted the coverage of DNA molecules without features. The resulting coverage profile shows a distinct peak at the *E. coli* origin of replication in the bubble-enriched sample, but was absent in the unreplicated control (**Figure 3.4**). This result is consistent with detection of the *E. coli* origin based on DNA fluorescence intensity.

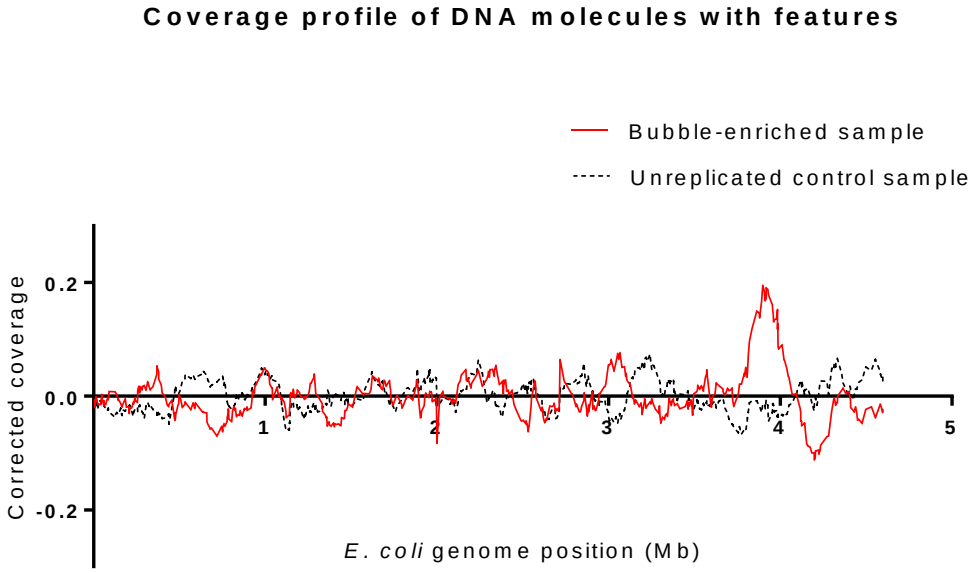


Figure 3.4. Coverage profile of DNA molecules with features.

The coverage profile is the coverage of DNA molecules with features, minus the coverage of DNA molecules without features. A distinct peak is located at the *E. coli* origin of replication (3.92Mb) (Blattner et al., 1997).

Single molecule DNA fluorescence intensity

To perform single molecule analysis, DNA molecules mapped to the *E. coli* origin were identified. Using a segmentation algorithm, the YOYO-1 intensities along each molecule were plotted as square waves (**Figures 3.5a and 3.6a**). The width of the square wave represents the length of the region of elevated fluorescence intensity. Thus, for replication bubbles, the width of the square wave is the size of the bubble.

In both the bubble-enriched sample and the unreplicated control sample, DNA molecules with feature(s) were mapped to the *E. coli* origin (**Figures 3.5 and 3.6a**). No filter was applied to screen out DNA molecules whose square wave widths were narrow. DNA molecules that had narrow square waves, and whose square waves were further away from the origin, are less likely to be true replication bubble.

Single DNA molecule nick labels

Next, we examined the putative nick labels on single DNA molecules. Some DNA molecules mapped to the origin had putative nick labels at the edges of the square wave (**Figure 3.5b, c**). At a *bona fide* replication bubble, the square wave edges represent the replication forks.

A replication fork contains a leading strand with a free 3'-hydroxyl group. The replication fork can be extended with a DNA polymerase in a primer extension-like manner. We speculated that the putative nick labels at the edge of the square waves do not represent nick sites. Rather, we reasoned that the labels at the edge of the square waves represent labels incorporated at replication forks (**Figure 3.7**). Thus, we

hypothesized that labeling the replication fork first in one color, then performing the nick translation step in a second color, may increase the sensitivity of detecting origins.

Figure 3.5. Single molecule DNA fluorescence intensity from replication bubble-enriched samples.

A) The YOYO-1 fluorescence intensity of DNA molecules that overlap the *E. coli* origin of replication (green line), plotted as fitted square-waves. **B and C)** Two specific DNA molecules with their YOYO-1 fluorescence intensity (squiggle) and computed square-waves are shown. Nick sites are shown as dots. Top row of dots: the location of presumptive nick labels detected on the DNA molecule. Bottom row of dots: the location of the nick labels in the reference genome. Purple dots indicate a positive match between the two rows; red dots indicate a negative match. Arrow indicates labels at the presumptive replication fork that were present on the DNA molecule but absent in the reference genome.

Figure on next page

Figure 3.5. Single molecule DNA fluorescence intensity from replication bubble-enriched samples.

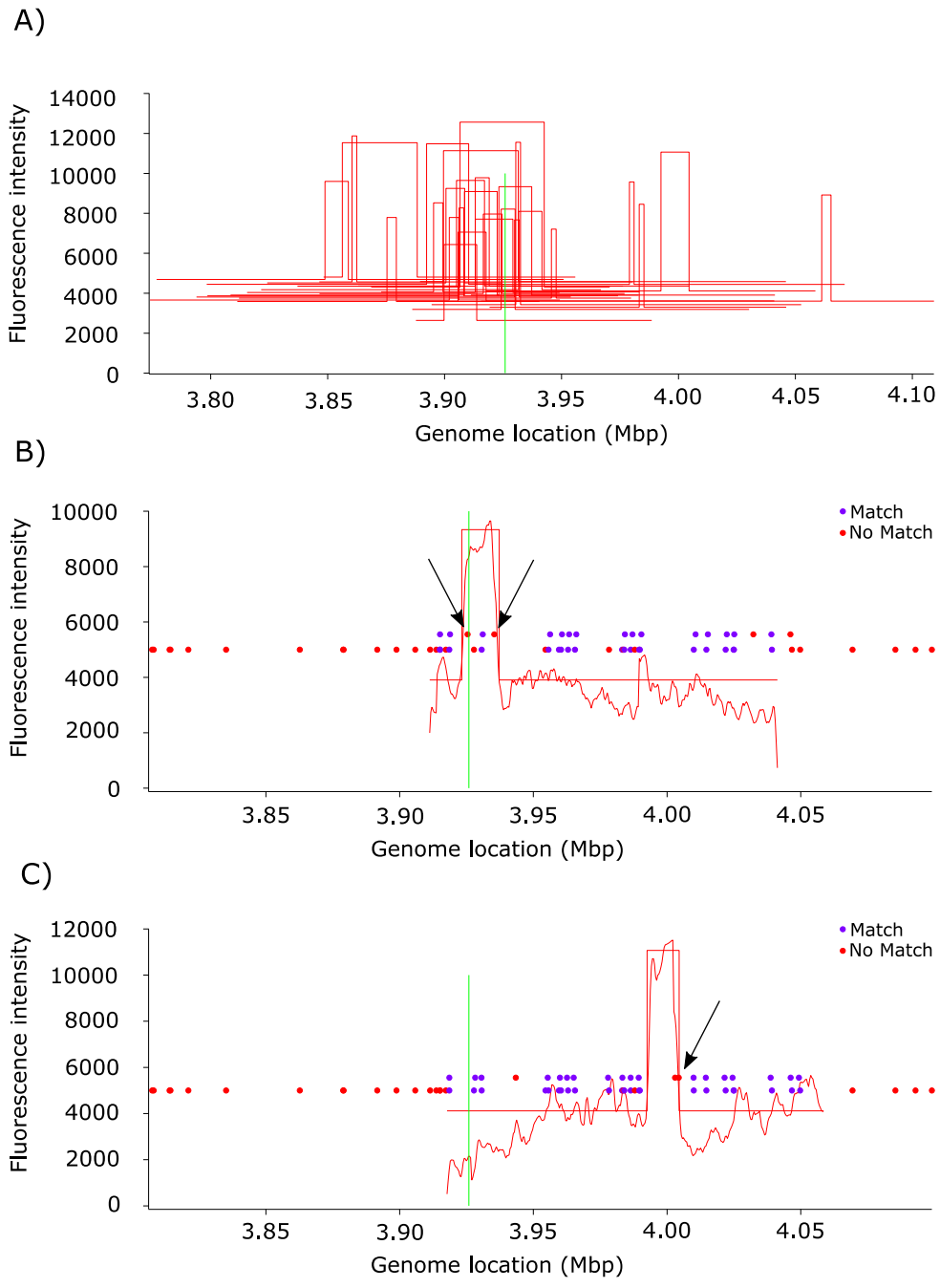
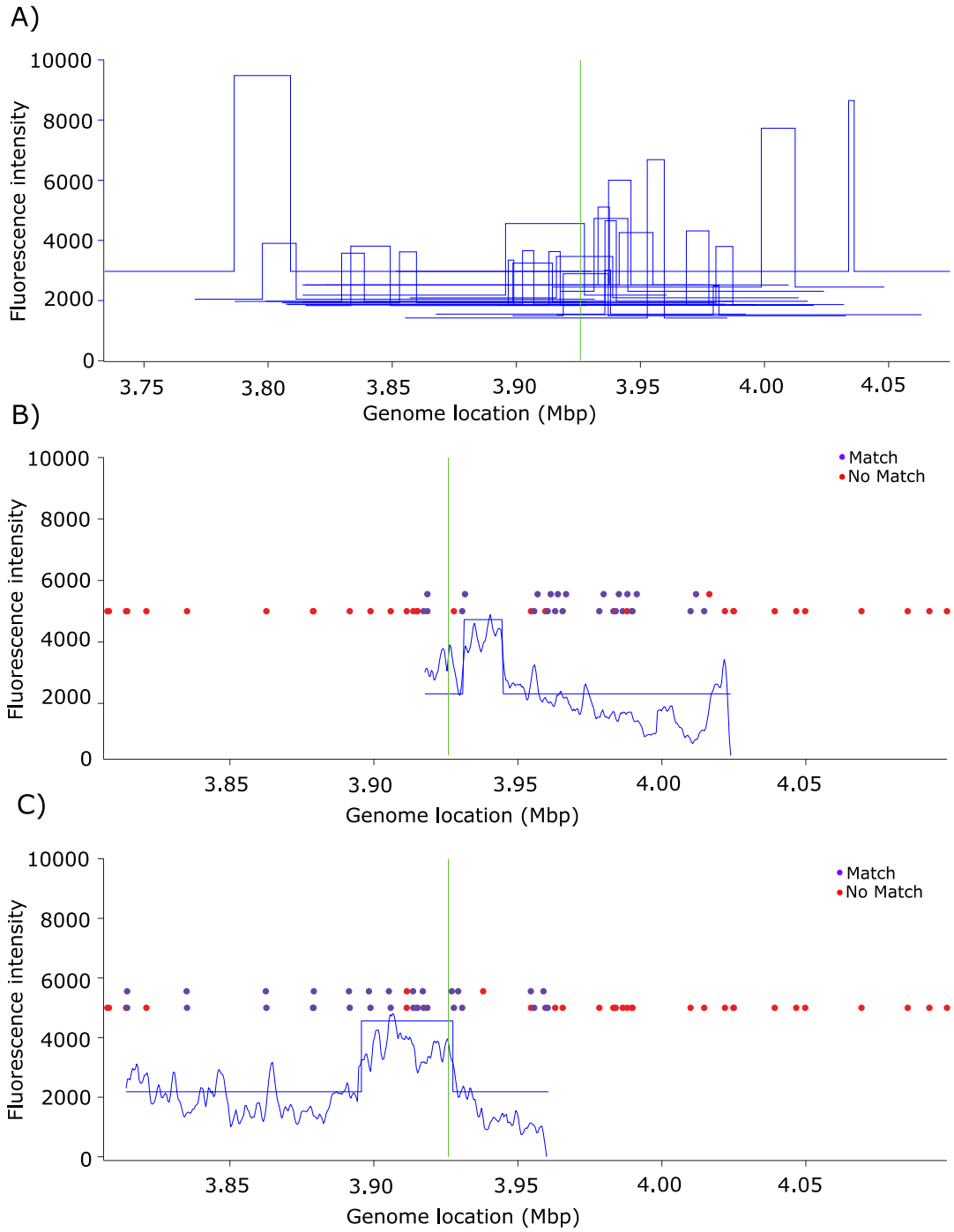


Figure 3.6. Single molecule DNA fluorescence intensity from unreplicated control samples.

A) The YOYO-1 fluorescence intensity of DNA molecules that overlap the *E. coli* origin of replication (green line), plotted as square-waves. **B and C)** Two specific DNA molecules with their YOYO-1 fluorescence intensity (squiggle) and computed square-waves are shown. Nick sites are shown as dots. Top row of dots: the location of presumptive nick labels detected on the DNA molecule. Bottom row of dots: the location of the nick labels in the reference genome. Purple dots indicate a positive match between the two rows; red dots indicate a negative match.

Figure on next page

Figure 3.6. Single molecule DNA fluorescence intensity from unreplicated control samples.



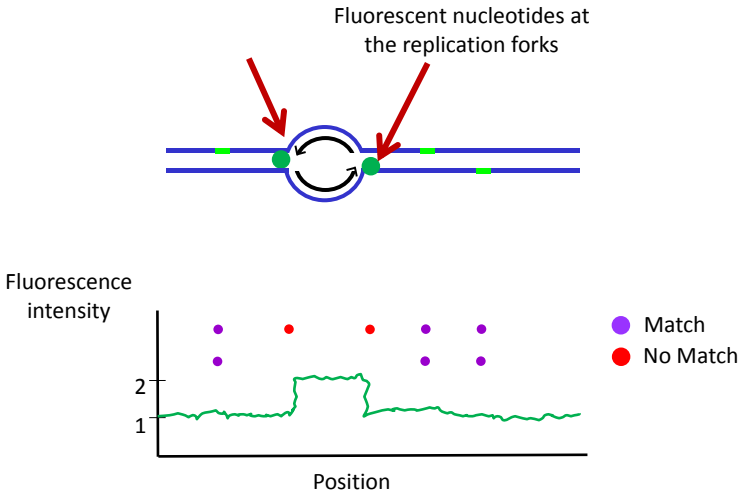


Figure 3.7. Interpretation of labels at the edge of the fluorescence intensity square-waves.

During the nick labeling step, the leading strands at a replication bubble can incorporate fluorescent nucleotides. The labels at the edge of the square-waves can thus denote the location of replication forks.

3.3.2. Strategy #2: Bubble labeling (replication fork labeling)

Labeling of replication forks in solution

Our second approach to detecting replication origins on single DNA molecules is to label replication forks. Here, we used a two-color labeling scheme: first, DNA molecules were treated with Taq DNA polymerase and red fluorescent nucleotides. This first step was designed to label replication forks. After a washout, DNA molecules were nick labeled with green fluorescent nucleotides. This second step was designed to provide nick labels for alignment.

After two-color labeling, DNA molecules were mapped to the *E. coli* reference genome. The location of incorporated red fluorescent nucleotides are marked as red labels. Each red label constitutes one or multiple red fluorescent nucleotides.

Around the origin, an enrichment of red labels was observed in the bubble-enriched sample (**Figure 3.8**). This enrichment was absent in the unreplicated control. This observation is consistent with replication forks being labeled.

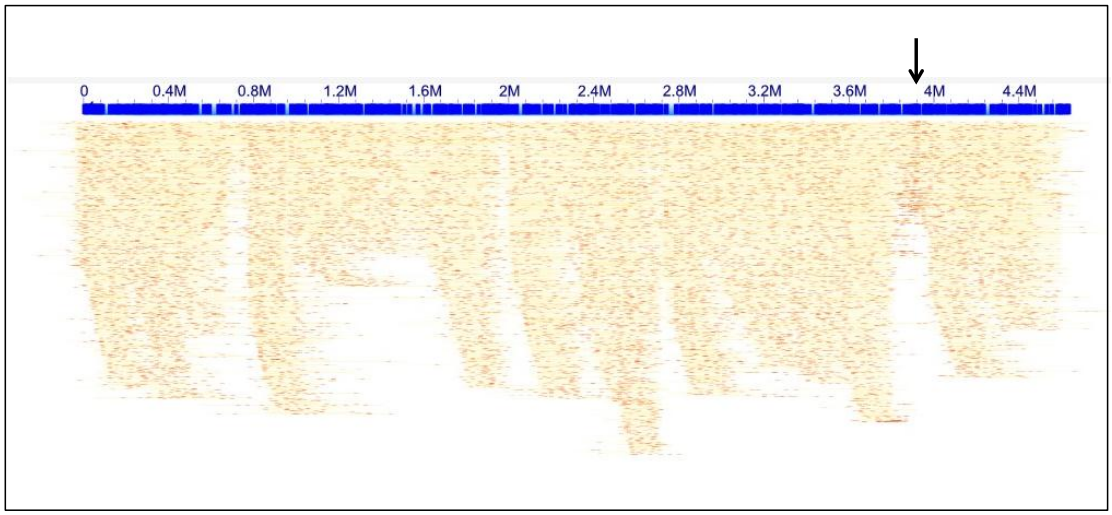
Figure 3.8. Two-color labeled DNA molecules mapped to the *E. coli* genome.

DNA molecules were first labeled in red (as shown) to mark the location of the replication forks. Next, DNA molecules were labeled in green to mark the location of known nick sites for genome mapping. An enrichment of red labels at the *E. coli* origin of replication (arrow) is present in the bubble-enriched sample, but absent in the unreplicated control. Image produced from BioNano Genomics IrysView software.

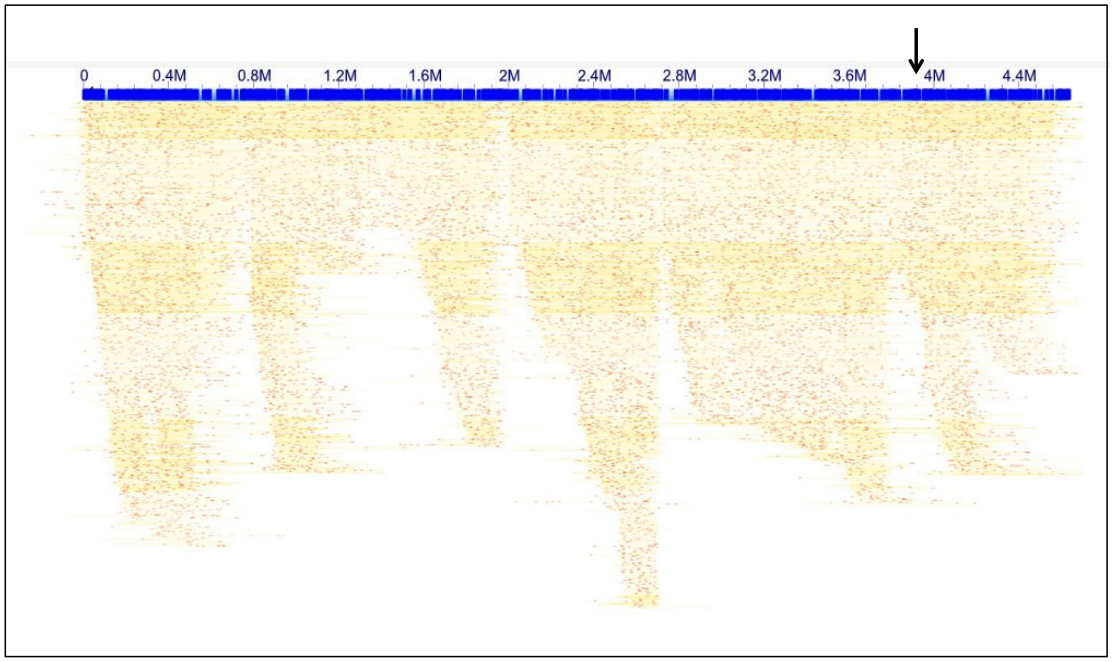
Figure on next page

Figure 3.8. Two-color labeled DNA molecules mapped to the *E. coli* genome

A) Bubble-enriched sample



B) Unreplicated control



Red labels were also present in DNA molecules across the genome. Genomic DNA isolation and handling can introduce non-specific DNA breaks and nicks. Red labels may thus be incorporated at both pre-existing nicks and at replication forks. Therefore, we investigated whether repairing non-specific nicks prior to two-color labeling would reduce the noise.

To repair non-specific nicks, we used DNA ligase. DNA molecules were then two-color labeled. With repair, the frequency of DNA molecules with red labels decreased. Among repaired DNA molecules mapped to the origin, 27% of molecules had 2 or more red labels. Among repaired DNA molecules that mapped outside of the origin, only 4% of molecules had 2 or more red labels (**Figure 3.9**). Thus, molecules mapped to the origin had 6-fold greater frequency of having 2 or more labels than non-origin molecules.

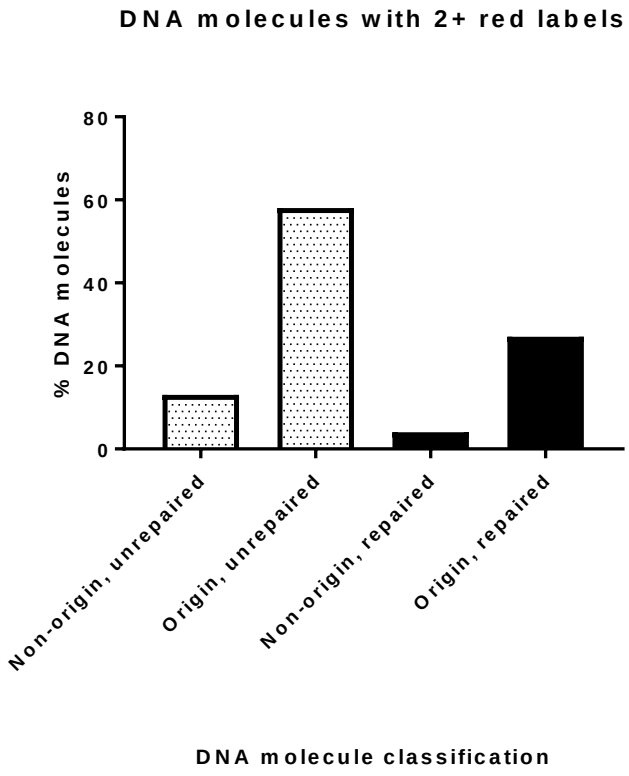


Figure 3.9. Percentage of DNA molecules that have 2 or more red labels.

DNA molecules from bubble-enriched samples were classified according to whether they were mapped to the *E. coli* origin of replication ("origin") or not ("non-origin"). Prior to labeling, DNA molecules were treated with DNA ligase ("repaired") or not ("unrepaired").

Genomic location of filtered red labels

To further decrease the noise from non-specific labeling during the replication fork labeling step, we added a filter to our analysis. Here, we selected for DNA molecules that had at least two red labels within a 60kb sliding window. This filter was designed to eliminate single outlier red labels that contributed to the non-specific background. We expect the replication forks in our bubble-enriched samples to be less than 60kb apart. This was determined based on the width of the square waves observed in single DNA molecules (**Figure 3.5a**). This filter selected for 140 of the 9864 DNA molecules (1.4%) with 2 or more red labels. With this filter, DNA molecules mapped to the genome with a distinct peak in coverage around the origin (**Figure 3.10**). This finding indicates that the replication forks emanating from the *E. coli* origin can be detected via our two-color labeling system.

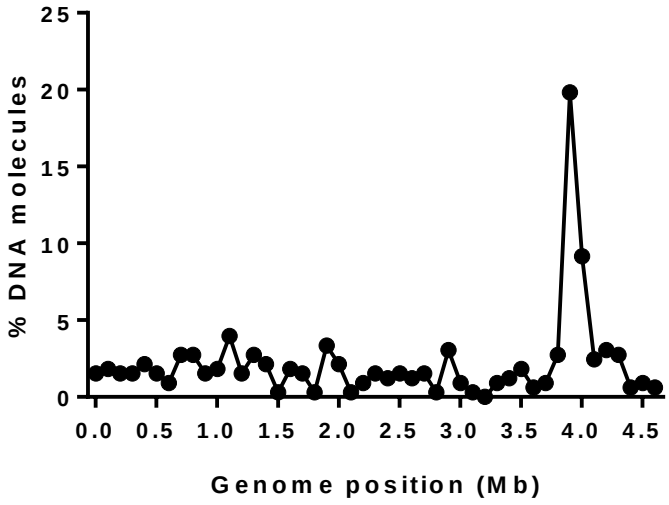


Figure 3.10. Frequency distribution of filtered DNA molecules with red labels.

DNA molecules with red labels that are within 60kb of each other on the same molecule were selected. This eliminates DNA molecules with single, isolated red labels. Filtered DNA molecules (n=328) were placed in 100kb bins. The distribution shows an enrichment around the *E. coli* origin (3.92 Mb). DNA molecules are from the bubble-enriched, repaired sample.

Labeling of replication forks in agarose plugs

In addition to labeling replication forks directly on genomic DNA in solution, we tested labeling replication forks on genomic DNA embedded in agarose plugs. Agarose stabilizes DNA and may increase the recovery of high molecular weight genomic DNA.

To extend the leading strand to a significant extent for fluorescence detection, a DNA polymerase must 1) have strand displacement activity to unravel the parental double-stranded DNA; and 2) incorporate fluorescent nucleotides. To confirm that Klenow (3' → 5' exo-) satisfies both requirements, I performed rolling circle amplification on a 7.2kb M13mp18 circular template in solution (**Figure 3.11a-c**). My results confirmed both requirements. Repeating this experiment with the M13mp18 circular template embedded in agarose plugs led to the same finding: fluorescence was detected in the amplified product, suggesting primer extension and incorporation of fluorescent nucleotides (**Figure 3.11d**).

Next, I asked whether Klenow (3' → 5' exo-) could extend the leading strands of replication forks in HEK293 genomic DNA embedded in agarose plugs. After the reaction, no fluorescence signal was detected in the genomic DNA (**Figure 3.11e**). For this experiment, asynchronous HEK293 cells were used. Since the majority of the cell cycle is outside of S-phase (i.e. most cells are not replicating and therefore do not have active replication forks), the fluorescence intensity may be below the threshold of detection. The use of synchronized cells to enrich for replication forks is a logical step

forward. However, since the labeling of replication forks in solution appeared to work, we did not pursue the approach of fork labeling in agarose plugs further.

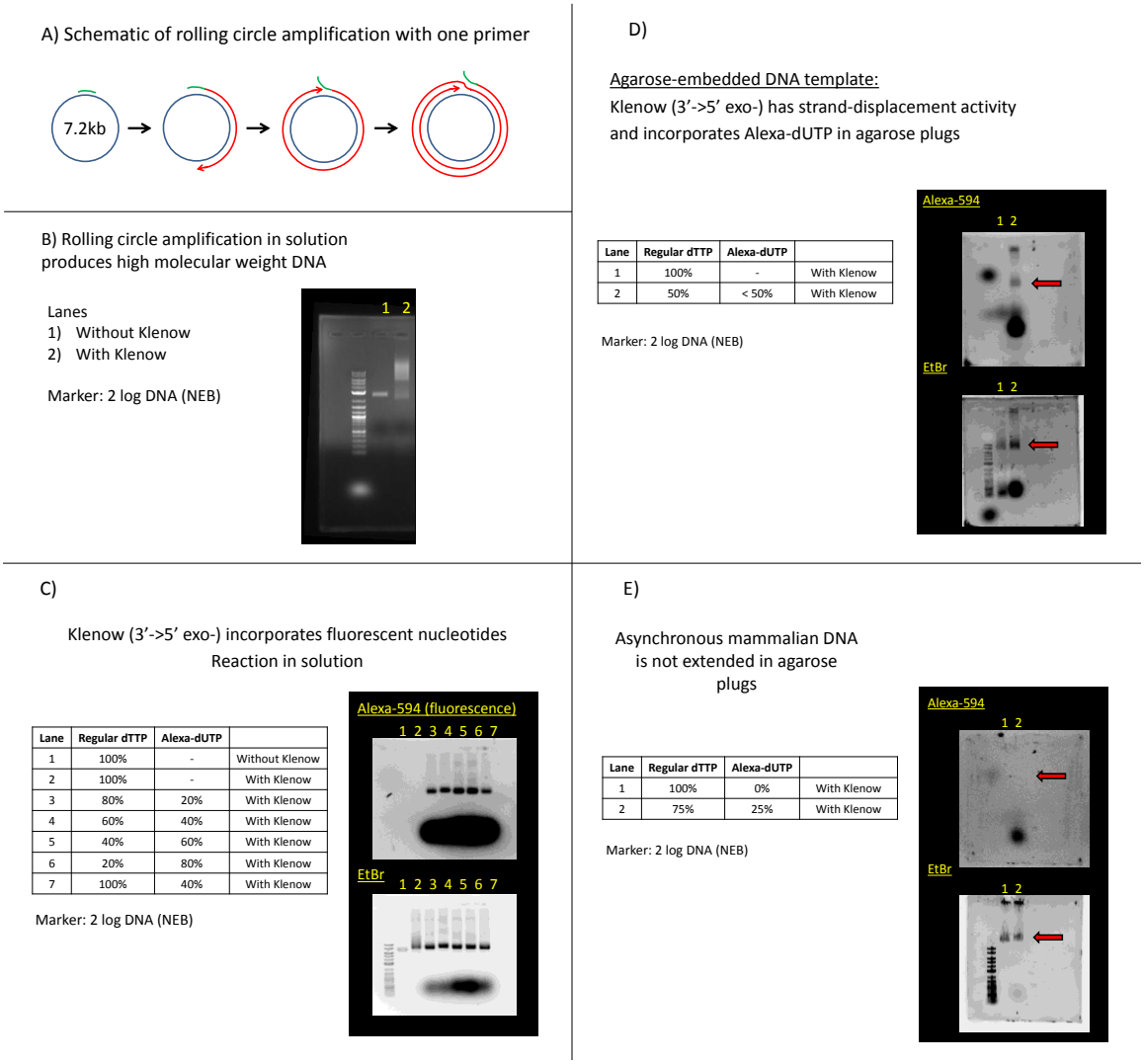


Figure 3.11. Rolling circle amplification and replication fork extension.

A) Schematic of rolling circle amplification with one primer. B) Rolling circle amplification with Klenow produces high molecular weight DNA. C) Klenow (3'→5' exo-) incorporates fluorescent nucleotides in a rolling circle assay in solution. D) Klenow (3'→5' exo-) incorporates fluorescent nucleotides in a rolling circle assay in agarose-embedded DNA templates. E) No evidence of replication fork elongation in asynchronous mammalian DNA embedded in agarose plugs.

3.3.3. Strategy #3: Nucleoside pulse labeling

Our third approach is to pulse label live cells with nucleoside analogs for optical detection in nanofluidic channels.

EdU detection in DNA molecules

To determine whether EdU labeling and detection is suitable for optical mapping of replication origins in nanofluidic channels, HeLa cells were treated with 10uM EdU overnight. EdU-labeled cells were detected using azide Alexa Fluor 594 via the CuAAC reaction. Whole cell staining confirmed nuclear detection of EdU (**Figure 3.12a**).

Optical mapping requires long DNA molecules for proper alignment. To determine whether DNA molecules are long enough for optical mapping, DNA molecules were stretched on silanized glass coverslips via molecular combing. DNA molecules after the CuAAC reaction were short. In contrast, untreated DNA molecules were long (**Figure 3.12b**). The short DNA fibers after CuAAC are problematic for optical mapping, because long DNA molecules are required for mapping nick sites.

The short DNA fibers after CuAAC are consistent with copper(I)-mediated DNA damage (**Figure 3.12c-d**). Copper(I) is thought to generate free hydroxyl radicals in the presence of molecular oxygen in a Fenton-like reaction (Urbañski and Beresewicz, 2000). We thus performed the CuAAC reaction in a nitrogen glove box—where the O_2 tension is low—but DNA shearing was still observed (**Figure 3.12e**).

Figure 3.12. Effects of EdU detection via CuAAC

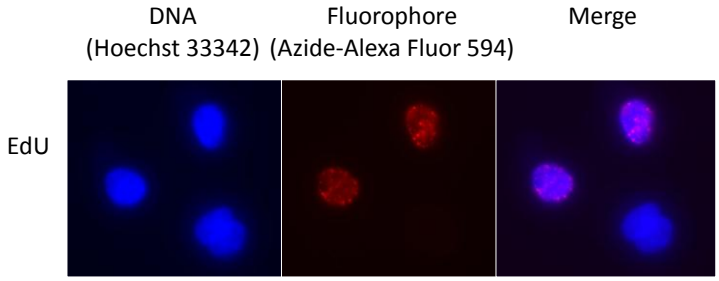
A) EdU staining is localized to the nucleus. HeLa cells were treated with 10uM EdU for overnight, harvested, and conjugated with Azide-Alexa Fluor 594 via CuAAC. The staining indicates EdU incorporation.

B) EdU-labeled DNA fibers are short after CuAAC. HeLa cells were treated with 25uM EdU or DMSO overnight, harvested, processed for fluorophore conjugation via copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) reaction. DNA was isolated, stained with YOYO-1, and stretched on glass coverslips via DNA combing. Width of frame represents 350kb. DNA fibers are short after CuAAC.

Figure on next page

Figure 3.12. Effects of EdU detection via CuAAC

A) EdU detection via CuAAC



B) DNA fibers via molecular combing

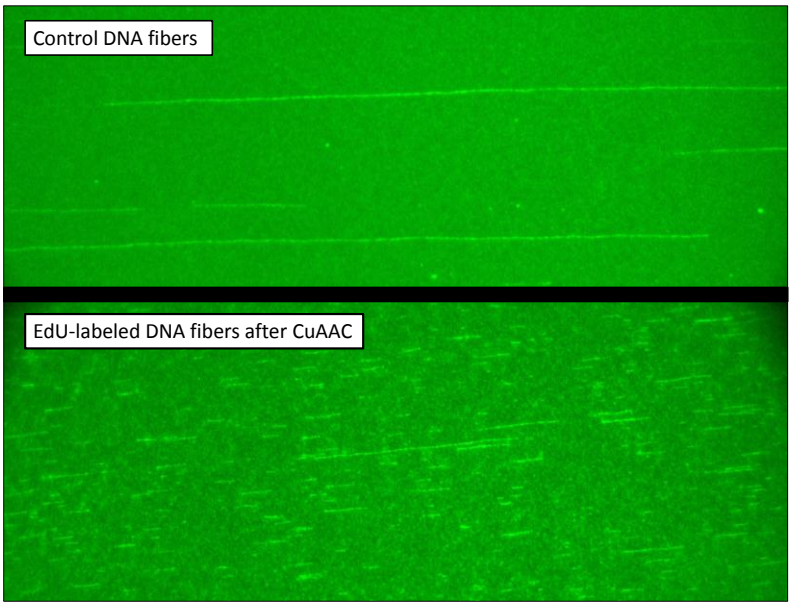


Figure 3.12. (continued)

C) DNA after CuAAC is smeared on pulsed field gel electrophoresis. HeLa cells labeled with EdU or no EdU overnight were harvested, and treated via CuAAC or control conditions. The copper in $CuSO_4$ is in the inactive form (Cu^{2+}), whereas in CuAAC is the active form (Cu^{1+}). The finding is indicative of copper(I)-mediated DNA damage.

D) Copper(I)-dependent DNA damage. Lambda DNA shows [Cu^{1+}]-dependent DNA smearing on gel electrophoresis. Ladder is the 2 log DNA ladder (New England Biolabs).

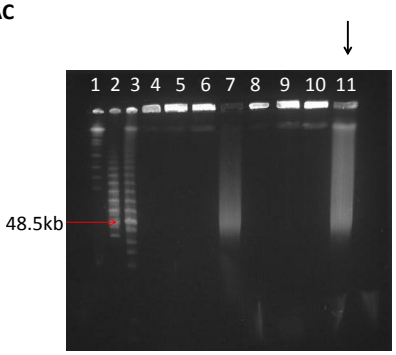
E) Nitrogen glove-box CuAAC. HeLa cells labeled with EdU were treated via CuAAC or control conditions in a nitrogen-filled glove box to provide an anaerobic environment. This approach still produced fragmented DNA.

Figure on next page

Figure 3.12. (continued)

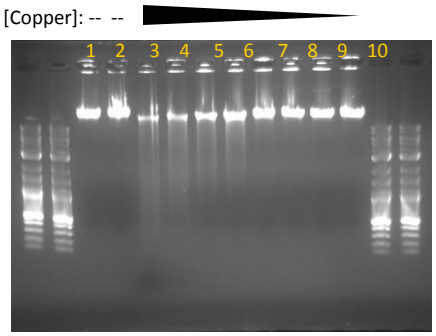
C) Pulsed field gel electrophoresis after CuAAC

Lanes
1) Yeast chromosome marker
2) Mid-range marker
3) Low-range marker
4) No EdU, no fixation.
5) No EdU, PBS treated
6) No EdU, $CuSO_4$ control (no CuAAC)
7) No EdU, CuAAC (click chemistry)
8) EdU, no fixation.
9) EdU, PBS treated
10) EdU, $CuSO_4$ control (no CuAAC)
11) EdU, CuAAC (click chemistry)



D) [Copper(I)]-dependent λ DNA damage

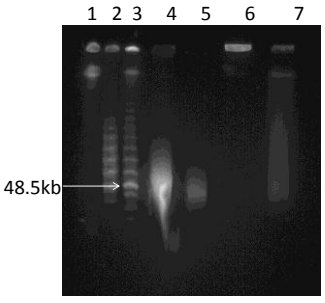
[Copper concentration]
in each lane:
1) 0mM
2) 0mM
3) 3.7mM
4) 1mM
5) 0.37mM
6) 0.1mM
7) 0.037mM
8) 0.01mM
9) 0.0037mM
10) 0.001mM



E) Nitrogen glove-box CuAAC

Lanes
1) Yeast chromosome marker
2) Mid-range marker
3) Low-range marker
4) EdU labeled, CuAAC (N_2)
5) EdU labeled, CuAAC (N_2)

Controls
6) EdU labeled, no CuAAC (air)
7) EdU labeled, CuAAC (air)



Copper coordinating ligands have been reported to increase CuAAC labeling efficiency and decrease copper(I)-mediated cytotoxicity (Bevilacqua et al., 2014; Jiang et al., 2014; Soriano del Amo et al., 2010; Wang et al., 2011; Zheng et al., 2013). We asked whether the ligand BTTP could decrease copper-mediated DNA damage yet allow EdU to still be detectable. Compared to $CuSO_4$, using a premix of BTTP and $CuSO_4$ (4:1 ratio) decreased lambda DNA damage. However, the fluorescence of EdU detection was significantly decreased (**Figure 3.13**). Thus, it is possible that the effective copper(I) concentration was decreased, leading to less DNA damage (desirable effect) but also lower CuAAC reaction efficiency (undesirable effect). Although it is possible that the lower CuAAC reaction efficiency is enough for optical mapping, we decided to pursue an approach that did not require EdU or a copper(I) catalyst.

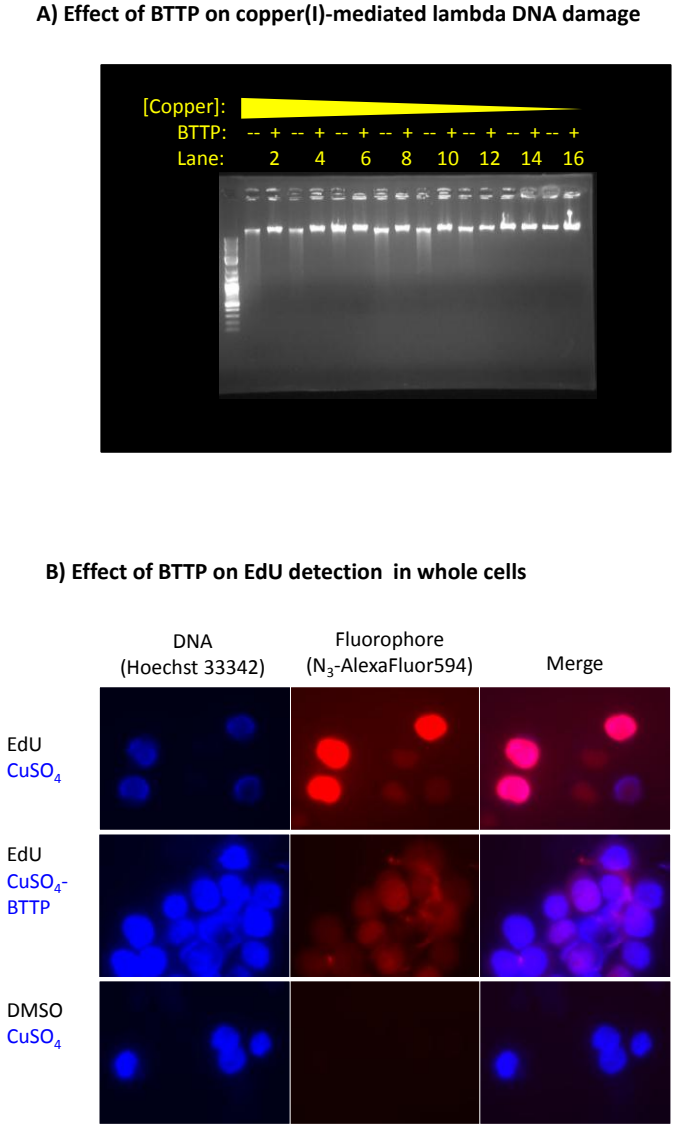
Figure 3.13. BTTP ligand lowers Cu(I)-mediated DNA damage, but also lowers EdU detection efficiency.

A) Gel electrophoresis of lambda DNA treated with varying concentrations of copper and BTTP. Lambda DNA shows [Cu^{1+}]-dependent DNA smearing on gel electrophoresis from $CuSO_4$ but not from $CuSO_4$-BTTP premix. Ladder is the 2 log DNA ladder (New England Biolabs). The following concentrations of $CuSO_4$ were used: Lanes 1 and 2 (1mM); Lanes 3 and 4 (0.37mM); Lanes 5 and 6 (0.1mM); Lanes 7 and 8 (0.037mM); Lanes 9 and 10 (0.01mM), Lanes 11 and 12 (0.0037mM); Lanes 13 and 14 (0.001mM); Lanes 15 and 16 (0mM).

B) Whole cell staining of EdU-labeled HeLa cells. CuAAC-mediated whole cell fluorescence is weaker with $CuSO_4$-BTTP than $CuSO_4$ alone.

Figure on next page

Figure 3.13. BTTP ligand lowers Cu(I)-mediated DNA damage, but also lowers EdU detection efficiency.



AmdU detection in DNA molecules

An alternative to EdU-labeling is to use AmdU, a thymidine analog which contains a reactive azide group. The azide group can participate in an azide-alkyne cycloaddition reaction. In the presence of a ring-strained alkyne, a copper catalyst is not required for the reaction. This approach eliminates any concern of copper-mediated DNA damage.

To investigate whether AmdU labeling can be used for optical mapping, we used MEL-HSV-tk cells (Feng et al., 2005), which have the promiscuous HSV thymidylate synthase gene to metabolize AmdU to its triphosphate form for DNA incorporation (Neef and Luedtke, 2014). AmdU can be detected in two ways: 1) CuAAC (with copper); or 2) SPAAC (without copper). First, we established that AmdU is detectable in the nucleus when treated with alkyne Alexa Fluor 594 via CuAAC. Next, we tested whether AmdU is detectable without copper. Both the ring-strained alkyne fluorophores, DIBO-Alexa Fluor 594 and DBCO-PEG4-TAMRA, showed non-specific diffuse whole cell staining (**Figure 3.14**). Thus whilst AmdU is incorporated, detection using ring-strained alkyne remains inconclusive as to whether the SPAAC reaction has occurred.

Figure 3.14.

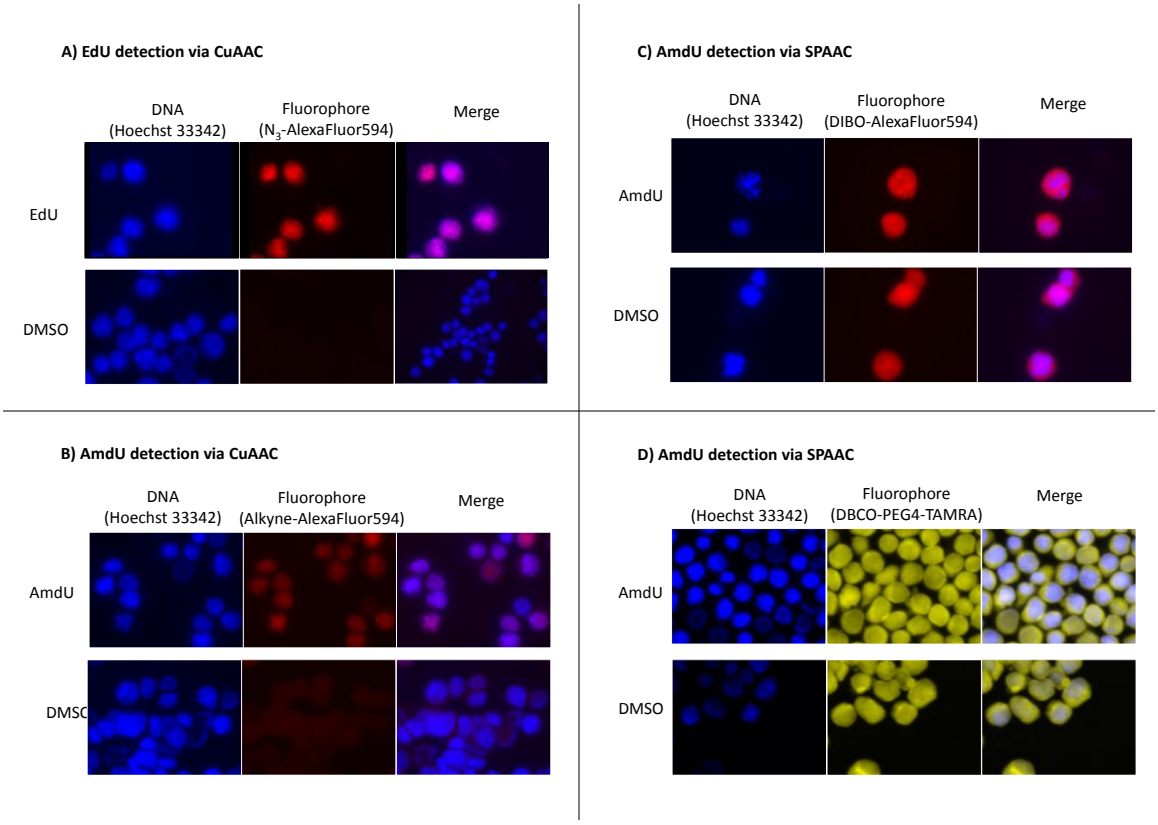


Figure 3.14. Detection of thymidine analogs in MEL-HSV-tk cells in whole cells.

Both **A)** EdU and **B)** AmdU are detected in the cell nucleus via CuAAC. However, AmdU detection via SPAAC is inconclusive, because both the **C)** DIBO Alexa Fluor 594 and **D)** DBCO-PEG4-TAMRA dyes show non-specific cellular staining.

Next, we examined AmdU-labeled DNA fibers treated with DIBO-Alexa Fluor 594. DNA fibers visualized via molecular combing showed no fluorescence signal. However, it should be noted that unlike immunofluorescence, where there is antibody signal amplification, the use of DIBO-Alexa Fluor 594 involves no amplification. Therefore, the absence of a visible signal does not dismiss the possibility that the SPAAC reaction has occurred.

Lastly, we synthesized PCR products containing AmdU at all thymidine positions. This was achieved in using AmdUTP in lieu of dTTP in the PCR cocktail. The AmdU-containing PCR products were labeled overnight with DBCO-OH or BCN-OH. The lower gel mobility indicates that conjugation has occurred (**Figure 3.15**).

Figure 3.15.

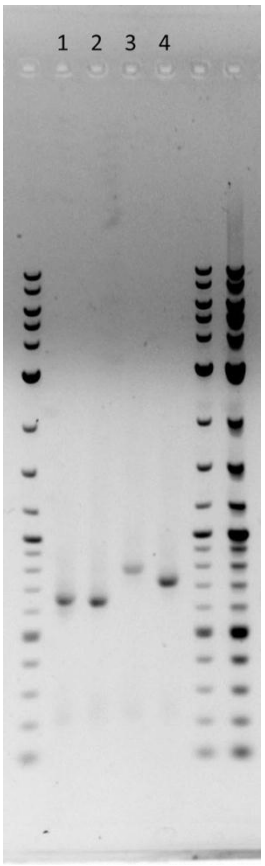


Figure 3.15. AmdU in PCR products can conjugate to DBCO-OH and BCN-OH.

The four sample lanes (left-to-right) represent PCR products with AmdU that are: 1) unlabeled; 2) unlabeled; 3) labeled with DBCO-OH; and 4) labeled with BCN-OH. PCR products were prepared using dATP, dCTP, dGTP, and AmdUTP (from Prof. Tom Brown, Oxford University). Labeling was performed overnight at room temperature. Ladder is 2 log DNA ladder (New England Biolabs).

3.4. Discussion

In this study, we investigated three approaches to optically detect DNA replication origins and forks on single DNA molecules in nanofluidic channels.

In the first approach, we stained DNA molecules with the YOYO-1, a DNA binding fluorescent dye and mapped them to the *E. coli* genome. We found that DNA molecules did not distribute evenly throughout the *E. coli* genome. This finding suggests difficulty in aligning DNA molecules to certain regions in the genome, and could be due to sparse nick labels. Future studies could involve using nicking endonucleases that recognize other sequences. We corrected for the uneven distribution and generated a coverage profile for DNA molecules with regions of >2-fold increase in YOYO-1 intensity. We found a distinct peak at the origin in our bubble-enriched sample that was absent in the unreplicated control. Our results suggest that DNA fluorescence detection is a plausible approach to mapping replication origins.

The unreplicated control should not have any DNA molecules with feature(s). However, DNA molecules with feature(s) were identified in the unreplicated control at a comparable frequency. Topological events in nanochannels, such as DNA folds and DNA knots, can account for higher relative DNA fluorescence intensity (**Figure 3.16**) (Reifenberger et al., 2015). DNA folds and DNA knots can be identified by the shortening of intervals between nick labels (Reifenberger et al., 2015). In this study, we did not filter out DNA folds and DNA knots. Additionally, two DNA molecules can co-translocate into a nanochannel, which can produce regions of double fluorescence

intensity. As the analysis pipeline continues to be refined, we can expect an increase in the specificity of origin detection.

Figure 3.16

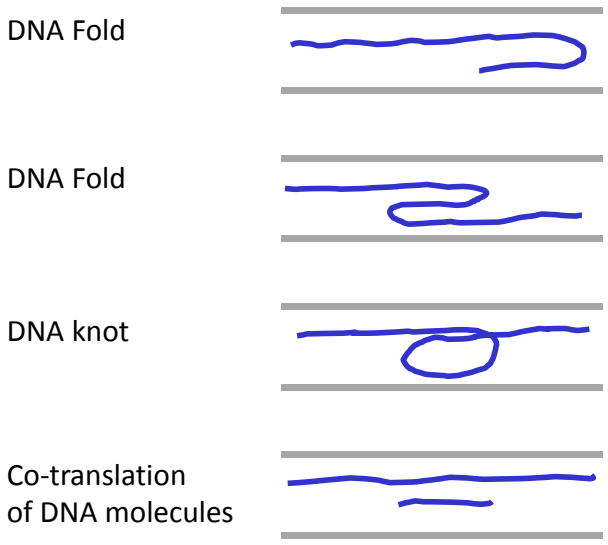


Figure 3.16. Topological and co-translational events may lead to increased fluorescence intensity.

DNA folds, DNA knots, and co-translation of DNA molecules in a nanochannel can produce regions of higher fluorescence intensity.

Our second approach to identify replication origins is a two-color sequential labeling approach: 1) labeling of replication forks via leading strand extension with red fluorescent nucleotides; and 2) labeling of nick sites with green fluorescent nucleotides. In our bubble-enriched *E. coli* sample, we observed an enrichment of red labels at the *E. coli* origin. However, many DNA molecules mapped throughout the genome also had red labels. Possible explanations include the presence of non-specific DNA breaks and nicks that were not repaired, the carryover of red fluorescent nucleotides from the first step (leading strand extension) to the second step (nick translation), and a leaky arrest that allowed replication forks to extend beyond the origin locus.

We extended our analysis and selected DNA molecules that had at least two red labels within a 60kb sliding window. In the bubble-enriched sample with DNA ligase pre-treatment, the filter selected for 140 molecules with 2 or more labels. A distinct peak at the origin was observed in a plot of the filtered red labels. Our results suggest that with filtering, labeling of replication forks is a plausible approach that can combine with DNA fluorescence analysis to map replication origins and replication forks. The combination of both the fork labeling approach and the DNA fluorescence intensity mapping approach raises the prospect of increasing the specificity of detecting *bona fide* origins.

Our third approach to identify replication origins on single DNA molecules is the pulse labeling of cells with nucleoside or nucleotide analogs. EdU detection via CuAAC produced short DNA molecules, making it difficult to generate enough nick labels per DNA molecule for genome mapping. We explored using AmdU, whose reactive azide group can react with a ring-strained alkyne in a copper-free cycloaddition reaction

(SPAAC). The ring-strained alkyne fluorophores that we tested resulted in diffuse background staining in whole cells. Therefore, it remains inconclusive—based on whole cell staining—whether the two fluorophores we used conjugated to AmdU. However, AmdU detection in whole cells has been demonstrated via conjugation with another dye, bicyclo[6.1.0]nonyne (BCN) Alexa Fluor 488 (Neef and Luedtke, 2014). AmdU detection using BCN-derived fluorophores have also been demonstrated in PCR products (Ren et al., 2015). BCN contains a ring-strained alkyne and can participate in SPAAC reaction (Dommerholt *et al.* 2010). AmdU in PCR products can be conjugated to two ring-strained alkynes, DIBO-OH and BCN-OH (**Figure 3.15**). Thus, AmdU detection with a ring-strained alkyne remains a highly plausible approach to detecting replication origins and forks. Another labeling strategy to be explored is the alkene-tetrazine ligation (Liu et al., 2012): the postsynthetic labeling of alkene-modified DNA with a tetrazine dye. This approach may include the use of 5-vinyl-2'deoxyuridine (VdU) conjugated to a tetrazine dye via an inverse electron demand Diels-Alder reaction (Rieder and Luedtke, 2014) or the detection of 7-vinyl-7-deaza-2'-deoxyadenosine with a tetrazine dye (Bußkamp et al., 2014). Additionally, VdU detection might be possible with the development of dyes containing triazolinedione (Naik et al., 2017).

In this study, we showed that the *E. coli* origin can be mapped using the BioNano Genomics Irys platform. Both the DNA fluorescence detection approach and the replication fork labeling approach identified the *E. coli* origin. The technology can be expanded to map replication origins and replication forks in other model organisms and mammalian cells.

Chapter IV: Nanopore detection of DNA replication sites

A Brief Communication

4.1. Nanopore sequencing

Nanopore sequencing is a real-time, next-generation DNA sequencing technology that does not require DNA template amplification (Kasianowicz et al., 1996). The nanopore sequencing approach involves the passage of DNA molecules in a flow cell with many pores ("nanopores") embedded on a platform or a membrane (Lu et al., 2016). The types of nanopores used have been biological or solid-state (Branton et al., 2008). The biological pores include the *Staphylococcus aureus* alpha-hemolysin pore, the *Mycobacterium smegmatis* porin A (MspA) pore, and the *E. coli* CsgG pore (Brown and Clarke, 2016; Butler et al., 2008; Carter and Hussain, 2017; Derrington et al., 2010; Goyal et al., 2014; Kasianowicz et al., 1996). The alpha-hemolysin pore is an extracellular protein secreted by the bacterium *Staphylococcus aureus*, and has a 1.5nm diameter through which DNA molecules can translocate (Gouaux, 1998; Kasianowicz et al., 1996; Song et al., 1996). The MspA pore has a comparable diameter of ~1nm (Deamer et al., 2016; Faller et al., 2004). These pore diameters allow single-stranded DNA to traverse, but are too narrow for the B-form of double-stranded DNA, which have a diameter of 2nm (Watson and Crick, 1953a). A constant voltage is applied across the platform or membrane, and the ionic current drives the DNA molecules through the pores. The electrical conductance is recorded across each nanopore. As a DNA molecule traverses across a nanopore, the electrical conductance across a nanopore decreases in a sequence-dependent manner. This signal is also dependent on several neighboring bases (Bayley, 2015). Thus, the nanopore functions as a sensor, and the signature change in the electrical conductance can be used to determine the DNA sequence (**Figure 4.1**). The

prospect of using nanopore sequencing is fascinating: the Oxford Nanopore Technologies MinION sequencer—which contains a flow cell array of nanopores—is a pocket-sized USB-powered device that has been used in the field such as the real-time sequencing of the Ebola virus during the outbreak in West Africa that began in 2014 (Quick et al., 2016).

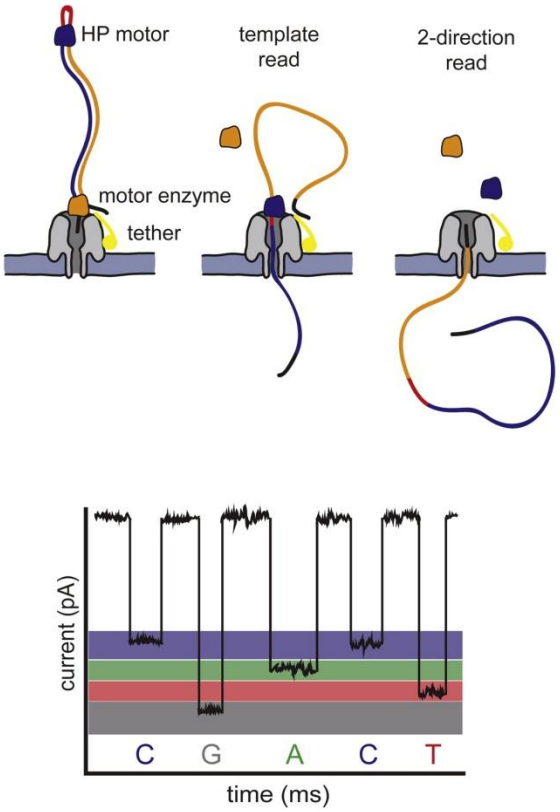


Figure 4.1. Principle of nanopore sequencing.

Double-stranded DNA is ligated to a hairpin (HP) motor on one end, and a motor enzyme on other end. The motor enzyme drives the translocation of DNA, as a single-strand, through the nanopore. The electrical blockade measured across the nanopore is used to determine the DNA sequence. Reprinted from Molecular Cell, 58, Reuter J. A., Spacek D. V., Snyder M. P., High-Throughput Sequencing Technologies, 586–597, Copyright (2015), with permission from Elsevier.

4.1.1. Nanopore detection of modified bases

Our laboratory was approached by our collaborators at Brown University (John Urban and Dr. Susan Gerbi), who hypothesized that nanopore sequencing technology can discriminate nucleoside analogs used in pulse labeling replication studies—such as BrdU and IdU—from the four canonical nucleosides. Consistent with this hypothesis, nanopore sequencing has been shown to be able to discriminate two bases that differ by a methyl group: 5-methylcytosine versus cytosine (Clarke et al., 2009; Ding et al., 2015; Laszlo et al., 2013; Manrao et al., 2011; Rand et al., 2017; Shim et al., 2013; Simpson et al., 2017; Wallace et al., 2010; Wescoe et al., 2014). Additionally, modified bases including 5-hydroxymethylcytosine, 5-carboxylcytosine, and 5-formylcytosine have also been discriminated from cytosine using nanopore sequencing (Laszlo et al., 2013; Wescoe et al., 2014; Zahid et al., 2016).

4.1.2. Approach to map replication sites

Based on this hypothesis, we embarked on a project to nanopore-sequence DNA with the modified bases BrdU, EdU, or IdU. Our strategy involves nanopore sequencing of three types of DNA: 1) PCR products with one or more types of modified bases; 2) *E. coli* genomic DNA with modified base labels at the origin of replication; and 3) *S. cerevisiae* genomic DNA with modified base labels at the early replicating origins.

The rationale for nanopore-sequencing PCR products is two-fold. First, it would provide a proof-of-principle that the modified thymidine analogs can be discriminated from thymidine. Second, the electrical readout from the known DNA sequence would be used to train a base-caller, such that an algorithm can be used to decipher an electric

readout into DNA sequences. The ability to base-call the four canonical bases has already been established, but the ability to base-call DNA molecules with BrdU, EdU, or IdU had not been established. Because our nanopore sequencing approach involves base-calling electrical readouts from 6-mers, we decided to synthesize PCR templates from a 6-mer de Bruijn graph, which has the sequence of all possible 6-mers from the four canonical bases. The minimum sequence of a 6-mer de Bruijn graph, composed of 4 bases, is 4096 bases (4^6 bases).

We first designed a 6-mer sequence that was 4.1kb. The sequence was synthesized as two inserts (2.68kb and 1.47kb) that were each cloned into pUC57 vectors. PCR amplification of the two inserts consistently produced partial length PCR products. An examination of the product size, and an analysis of the DNA sequence led me to postulate that the DNA polymerase was impeded by a high GC rich region. Indeed, the GC content around this region in one fragment was as high as 90%. Dimethylsulfoxide (DMSO) and betaine (a zwitterion at neutral pH) are additives that can improve PCR efficiency in templates with high GC content (Henke et al., 1997; Jensen et al., 2010; Rees et al., 1993; Varadaraj and Skinner, 1994). Based on this, I added DMSO or betaine to my PCR cocktail mixes, which increased the PCR product size, but it was still short of the full-length product. Next, John Urban designed a new 6-mer de Bruijn graph that had all possible 6-mer sequences, but in lower GC context. The four fragments (~1.3kb each) were cloned into pUC57 vectors, and the plasmids were used to generate PCR products.

To multiplex the PCR products generated from different analogs into one flow cell, we proposed to design PCR primers that had unique barcode regions at the 5'-

terminus. To ensure that the barcode region is recognized by the base-caller (which would allow us to determine the identity of the PCR product), I proposed that the barcode region of the primer be comprised of only three of the four canonical bases: C, G, and T. This so-called CGT-barcode would ensure that its complement would comprise of only G, C, and A. Therefore, the double-stranded barcode identifier would be devoid of any thymidine analogs. In designing the CGT-barcode sequences, we had to balance between having higher sequence complexity (to make each identifier distinctly unique) and having lower GC content (to decrease the likelihood of forming homodimers and heterodimers—that is, primer-dimers). Prior to selecting the barcode identifier sequences, each sequence was also screened to decrease the likelihood of off-target priming.

4.1.3. Preparation of analog-labeled PCR products

Two strategies were pursued to produce PCR products that had nucleoside analogs (BrdU, EdU, or IdU) substituted at the thymidine positions. The first approach (proposed by J. Urban, S. Gerbi, and N. Rhind) is a "hemi-labeling" approach, in which the PCR product generated from the four canonical dNTP's (dATP, dCTP, dGTP, and dTTP) were purified, denatured, and used as the template for one round of primer extension in the presence of a nucleotide analog (e.g. BrdUTP) in lieu of dTTP. This should result in double-stranded DNA in which one strand is labeled with BrdU. To enrich for hemi-labeled DNA, a forward 5'-biotinylated primers was used for primer extension. After primer extension, the DNA was enriched with streptavidin beads and released upon HindIII restriction digest. This approach, while plausible, produced hemi-labeled DNA that had low yield (**Figure 4.2**). It is possible that further optimization of

the stoichiometric ratio between the primer and PCR product concentrations may lead to higher yield.

Figure 4.2

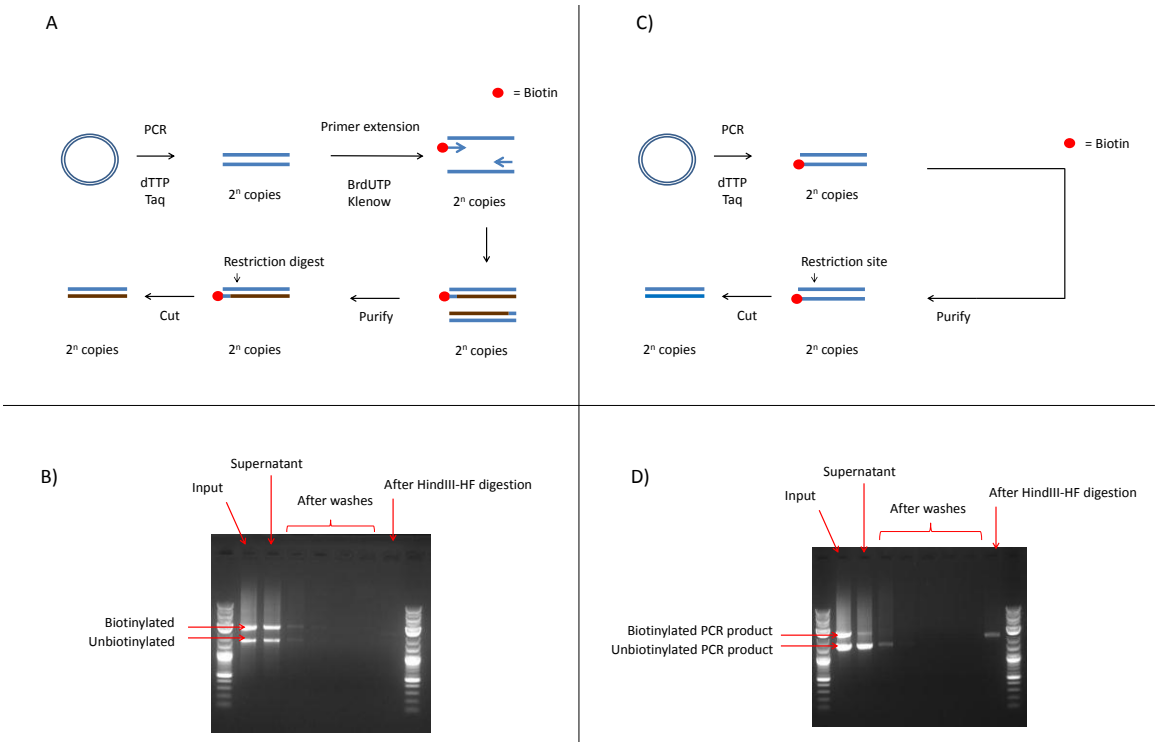


Figure 4.2. Hemi-labeling approach with biotin-streptavidin purification.

A) Schematic of hemi-labeling approach. PCR with the four canonical dNTP's (dATP, dCTP, dGTP, and dTTP) was followed by a primer extension step with a biotinylated primer and an unbiotinylated primer in the presence of dATP, dCTP, dGTP and BrdUTP. The product was purified on streptavidin beads and released upon HindIII digest.

B) Biotin-streptavidin purification produces low recovery with biotinylated primer extension. The biotinylated primer extension product and unbiotintylated control were incubated with streptavidin beads. After incubation, the supernatant was removed, the beads were washed 4 times, and the beads were restriction digested with HindIII-HF. Aliquots after each step were subjected to gel electrophoresis. Primer extension was performed with canonical dNTPs (dATP, dCTP, dGTP, and dTTP).

C) Schematic of biotin-streptavidin purification on biotinylated PCR product. Biotinylated PCR products were used to assess the functionality of the biotin-streptavidin purification.

D) Biotin-streptavidin purification produces high recovery with biotinylated PCR products.

Concurrently, I advocated a "full labeling" approach in which PCR products were generated from three canonical dNTP's (dATP, dCTP, and dGTP) and one nucleoside analog (e.g. BrdUTP). The advantages of this approach over the "hemi-labeling" approach are two-fold: 1) the hemi-labeling approach involves dTTP during PCR, which may be carried over to the primer extension step. The full labeling approach does not involve any dTTP, and thus would eliminate the concern for contaminants. 2) Fully labeled DNA would produce a better comparison to unlabeled control, because the DNA would not contain a mix of labeled and unlabeled DNA during the DNA preparation would eliminate any concern for any dTTP carried over to the primer extension step.

To develop the full labeling approach, I conducted PCR using three different polymerases (Deep Vent exo- DNA polymerase; Taq-T DNA polymerase; and Pwo DNA polymerase). While commonly used in standard PCR protocols, Taq DNA polymerase was unable to incorporate any of the nucleotide analogs we tested (BrdUTP, EdUTP, or IdUTP). After much troubleshooting with different polymerases, different thermocycler settings, and different reagent concentrations, I was able to generate full length PCR products with either BrdU, EdU, or IdU at all thymidine positions. These fully substituted PCR products were provided to John Urban, who used them to prepare libraries for sequencing on the Oxford Nanopore Techologies MinION sequencer. For these studies, the R9 verison of the MinION, which uses the *E. coli* CsgG pore, was used (Brown and Clarke, 2016). We also nanopore-sequenced PCR products prepared using barcode identifiers composed of the four canonical bases ("ACGT-barcodes"). The results showed more reads with the ACGT-barcode identifiers than with the CGT-barcode identifiers.

We have preliminary data from nanopore-sequencing of PCR products with full substitutions of BrdU, EdU, and IdU at thymidine positions (data not shown).

4.1.4. Nanopore sequencing genomic DNA containing 5 different bases

Our first step in base-caller training is to nanopore-sequence DNA where a specific nucleoside analog is substituted at all T positions. To do this, we have generated PCR products that contain all possible 6-mers formed from 4 different bases: A, C, G, and X, where X = BrdU, EdU, IdU, or T. However, due to the presence of endogenous T in cells, *in vivo* pulse labeling with one nucleoside analog would produce genomic DNA with at 5 different bases: A, C, G, T, and BrdU or EdU or IdU. Therefore, it is necessary to recognize the nucleoside analog in DNA composed of 5 different bases and not 4 different bases.

Two approaches have been proposed to detect the nucleoside analog in DNA containing 5 different bases. The first approach is to conduct base-caller training on DNA that contains all possible 6-mers formed from 5 different bases. This approach would require a de Bruijn graph of at least 5^6 = 15,625 bases. To generate the 15.6kb de Bruijn graph consisting of 5 different bases, it is possible to custom synthesize many oligonucleotides (e.g. ~156 oligonucleotides, with each oligonucleotide being a 100-mer). However, this approach is costly.

The second approach is to train the base-caller on the 6-mer de Bruijn graph from 4 different bases—that is, on PCR products that contain only {A, C, G, and T}, and separately, {A, C, G, and BrdU}. Signals from these libraries can be used to identify two

classes of 6-mers in the *in vivo* BrdU-pulse labeled DNA: those that contain T but no BrdU (for example, CGTTCC), and 6-mers that contain BrdU but no T (for example, CGBrBrCC). Regions with a third class of 6-mers—those that have a mixture of BrdU and T (for example, CGBrCGT)—might not be properly basecalled, but their genomic location may still be inferred based on abberant electrical blockade.

At present, we are pursuing the second approach, but the two approaches are not mutually exclusive and could potentially be pursued in parallel.

4.1.5. Nanopore sequencing-by-synthesis approach

The classic nanopore sequencing approach detects the electrical blockade from the translocation of single-stranded DNA through a nanopore. The standard library preparation involves taking double-stranded DNA and attaching a Y-shaped adapter on one end, and a hairpin adapter on another end (Lu et al., 2016). The Y-shaped adapter provides an opening to facilitate the entry of single-stranded DNA flap into the nanopore, and the subsequent unwinding of the duplex DNA template. The hairpin adapter forms a loop, such that a double-stranded DNA template can be read on both strands in one passage across a nanopore. Thus, the DNA is effectively sequenced twice. Other combinations can also be formed. For example, two hairpins would lead to a closed, dumb-bell shaped structure that does not provide a free end for entry into the nanopore. Attachment of two Y-shaped adapters to a double-stranded DNA molecule can lead to the independent sequencing of each DNA strand.

Recently, a sequencing-by-synthesis approach has been developed, in which the phi29 DNA polymerase is covalently coupled to alpha-hemolysin protein nanopores (Fuller et al., 2016; Kumar et al., 2012; Stranges et al., 2016). This approach utilizes tagged nucleotides that are incorporated by the anchored DNA polymerase. The newly synthesized DNA strand with the tagged nucleotides is translocated into the nanopore, and the DNA sequence is determined by the electrical blockade (**Figure 4.3**). This approach draws similarity to the PacBio SMRT sequencing technology discussed in Chapter II. Moreover, the PacBio RS II SMRT-sequencing platform with P6-C4 chemistry also uses phi29 DNA polymerase (Eid et al., 2009). There is recognized interest in combining nanopore with ZMWs (Larkin et al., 2014). As the nanopore sequencing-by-synthesis technology continues to develop, it would be fascinating to see whether the kinetics of tagged nucleotide incorporation can be resolved at the single-base level. Given that the kinetic parameters—interpulse duration and pulse width—observed in SMRT sequencing is a partly dependent on the DNA polymerase, it is conceivable that nanopore sequencing can be developed to measure the kinetics of nucleotide incorporation. Consequently, it is possible that the nanopore sequencing-by-synthesis approach can be developed to detect modified bases such as 5-methylcytosine and 5-hydroxymethylcytosine, and pulse labeled nucleoside analogs in DNA templates such as BrdU, EdU, IdU, and AmdU.

Figure 4.3.

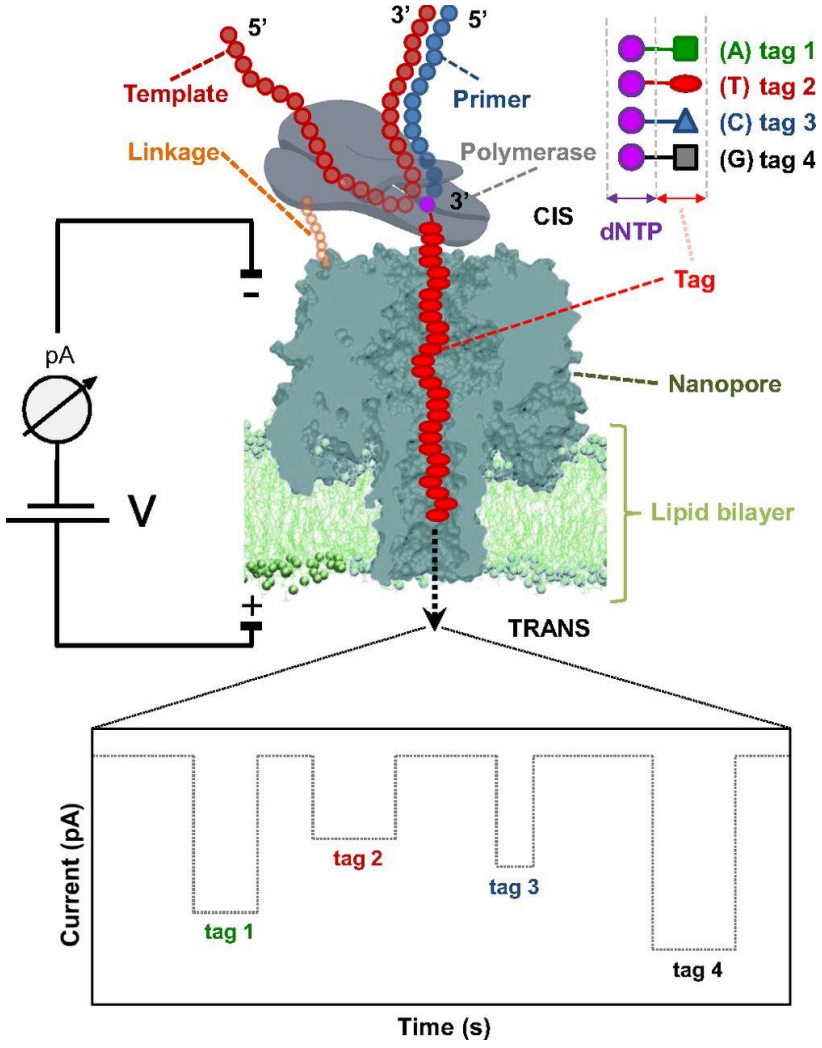


Figure 4.3. Nanopore sequencing-by-synthesis approach.

DNA polymerase is covalently attached to the alpha-hemolysin protein nanopore. Tagged nucleotides are incorporated along the DNA template, and the newly replicated DNA is translocated into the nanopore. The electrical blockade is used to determine the identity of the tagged nucleotide, and hence the DNA sequence of the original DNA template.

Figure is reproduced from Stranges P. B., et al., 2016, Design and characterization of a nanopore-coupled polymerase for single-molecule DNA sequencing by synthesis on an electrode array. *Proc. Natl. Acad. Sci.* 113: E6749–E6756.

Nanopore sequencing with both electrical and optical detection

The similarities between the nanopore sequencing-by-synthesis approach and the PacBio SMRT sequencing approach—both of which are compatible with phi29 DNA polymerase—opens the possibility of combining both detection technologies. In principle, the tagged nucleotides in the nanopore sequencing-by-synthesis approach can be fluorescently labeled, such that its incorporation can optically detectable. This can reduce the error rate of the current systems. Conversely, the zero-mode waveguides (ZMWs) of the SMRT sequencing system may be engineered to provide an electrical readout of nucleotide incorporation events.

The classic nanopore sequencing approach can also be engineered to allow simultaneous electrical and optical detection (McNally et al., 2010). This has been demonstrated with ~4nm solid-state nanopores, wherein electrical detection is combined with optical detection via total internal reflection fluorescence (TIRF) microscopy to monitor the translocation fluorescent double-stranded DNA and protein-DNA complexes (Soni et al., 2010). There is ongoing work to develop existing platforms that can simultaneously allow both electrical and optical detection (Liu et al., 2015, 2014). As the technology continues to develop, it will be fascinating to see whether pulse labeled DNA can be directly sequenced with higher accuracy in parallel sequencing arrays.

4.2. Conclusion

Nanopore sequencing of pulse-labeled DNA is an exciting real-time, single molecule approach to investigating DNA replication. The ability to detect multiple analogs via nanopore sequencing can facilitate the detection of replication origins, replication forks, and provide information on replication density, replication fork speed, and replication direction. Our work has provided an important foundation for which further development can lead to the ability to map replication sites on single DNA molecules at a high throughout rate. Furthermore, the possibility of applying nanopore sequencing to identify protein-DNA complexes on individual DNA molecules can advance our understanding on the mechanisms of DNA replication initiation.

4.3. Materials and Methods

A list of putative barcode identifier sequences was provided by John Urban (Brown University). The sequences were screened to minimize the likelihood of primer homodimerization, heterodimerization, and off-target priming. The 6-mer de Bruijn graphs (John Urban) was custom synthesized (GenScript, Piscataway, NJ) and cloned into pUC57 vectors.

Generation of hemi-labeled DNA: PCR was conducted with one 5'-biotinylated primer (VL-152), one unbiotinylated primer (MM24), and four canonical dNTP's (dATP, dCTP, dGTP, and dTTP) with Taq polymerase according to standard protocol. The plasmid pMM2 was as the template. PCR was performed in 50uL reactions using the following reagents: 1uL DNA template, 0.5uM forward primer, 0.5uM reverse primer, 0.2mM of each dNTP (dATP, dCTP, dGTP and dTTP), 1x Taq buffer, and 1uL Taq DNA polymerase (New England Biolabs).

Touchdown PCR was performed with the cycling conditions: 95^{o}C for 3mins, then 30 cycles (95^{o}C for 30secs, 65^{o}C for 45secs (-0.5^{o}C per cycle), 72^{o}C for 3mins), then 5 cycles (95^{o}C for 30secs, 50^{o}C for 45secs, 72^{o}C for 3mins), then 72^{o}C for 5mins, then 4^{o}C for 20mins.

Generation of fully labeled DNA: PCR was performed using barcoded primers that bind to sites flanking the insert in the plasmid templates. The plasmids containing portions of the 6-mer de Bruijn graph were pFS474, pFS475, pFS476, and pFS478. PCR was performed in 50uL reactions using the following reagents: 1ng plasmid DNA,

62.5nM forward primer, 62.5nM reverse primer, 1x Pwo SuperYield Buffer with Mg^{2+}, 0.2mM of each deoxyribonucleoside triphosphate (dATP, dCTP, dGTP, dXTP, where dX = BrdU, EdU, IdU, T, or a 1:1 mixture of the following pairs: BrdU & T, EdU & T, or IdU & T), and Pwo DNA polymerase (Roche).

Touchdown PCR was performed with the cycling conditions: 95^{o}C for 3mins, then 30 cycles (95^{o}C for 30secs, 65^{o}C for 45secs (-0.5^{o}C per cycle), 72^{o}C for 2mins), then 5 cycles (95^{o}C for 30secs, 50^{o}C for 45secs, 72^{o}C for 2mins), then 72^{o}C for 5mins, then 4^{o}C for 20mins.

Barcoded PCR products were provided to John Urban (Brown University) for library preparation and nanopore sequencing using the Oxford Nanopore Technologies MinION system.

Chapter V: Appendix – Identification of S-phase checkpoint targets in *Schizosaccharomyces pombe* using a flow cytometry based assay

5.1. Introduction

DNA replication is a fundamental biological process and—in eukaryotes—occurs in the S-phase of the cell cycle. Cells are frequently subjected to genotoxic and environmental stresses (Willis and Rhind, 2009b). Monitoring and responding to DNA damage insults in S-phase is critical, because genetic perturbations may alter the daughter cell genotype and lead to carcinogenesis in metazoans (Hoeijmakers, 2009). A surveillance mechanism that monitors and reacts to DNA damage during eukaryotic DNA replication is the S-phase checkpoint (Rhind and Russell, 1998). The S-phase checkpoint is a signaling cascade that is activated by DNA damaging insults such as ultraviolet irradiation (Rhind and Russell, 1998) and methyl methanesulfonate (MMS)-induced alkylation (Paulovich and Hartwell, 1995). MMS is a chemical mutagen that alkylates purines (Beranek, 1990). The S-phase checkpoint pathway is conserved in eukaryotes, and consists of three groups of proteins: sensor protein kinases, transducers, and effector kinases. Despite the importance of the S-phase checkpoint, the downstream targets of this pathway are not well understood. This chapter focuses on identifying the targets of the S-phase checkpoint effector kinase in the fission yeast *Schizosaccharomyces pombe*.

In *S. pombe*, the S-phase checkpoint response is initiated by the Rad3 protein, a member of the phosphatidylinositol kinase (PIK)-related family (Bentley et al., 1996; Cimprich et al., 1996; Keith and Schreiber, 1995; Seaton et al., 1992). Rad3 functions as the sensor kinase to detect DNA damage in S-phase. Rad3 is a homolog of the

mammalian ATM/ATR checkpoint proteins and the *Saccharomyces cerevisiae* Mec1 checkpoint protein (Bentley et al., 1996). The checkpoint signal is then transduced via Mrc1 to activate the S-phase checkpoint effector kinase, Cds1 (Lindsay et al., 1998; Tanaka and Russell, 2001). Cds1, a serine/threonine kinase, is a homolog of the mammalian Chk2 protein and *S. cerevisiae* Rad53 protein (Allen et al., 1994; Matsuoka et al., 1998; Sanchez et al., 1996; Sun et al., 1996; Weinert et al., 1994). Both Rad3 and Cds1 are required for the S-phase checkpoint response to UV-irradiation (Rhind and Russell, 1998) and MMS-induced alkylation in *S. pombe* (Kumar and Huberman, 2004; Marchetti et al., 2002).

The S-phase checkpoint responds to MMS-induced DNA alkylation damage by reducing the rate of bulk DNA replication (Paulovich et al., 1997; Paulovich and Hartwell, 1995). This checkpoint response is observable via flow cytometry, in which the DNA content is monitored throughout S-phase. In this assay, cells are synchronized (e.g. in the G1 phase) and released into S-phase in the presence of MMS (Willis and Rhind, 2011). Cells are then allowed to progress through S-phase, and samples are taken at timed intervals to determine the DNA content via flow cytometry. The DNA content at successive timepoints reflects the rate of bulk replication, which is reduced upon MMS treatment in *S. pombe* cells (Kumar and Huberman, 2004; Marchetti et al., 2002) and *S. cerevisiae* cells (Paulovich et al., 1997; Paulovich and Hartwell, 1995) that are checkpoint-proficient. Additionally, this assay can be used to determine whether strains in MMS are able to complete replication (like the checkpoint deficient *rad3*Δ strain or

*cds1*Δ strain) or unable to complete replication (e.g. wildtype strain at t=180 minutes after S-phase release) (Lindsay et al., 1998; Marchetti et al., 2002).

Mechanistically, any reduction in the bulk DNA replication rate may be due to the inhibition of origin firing, the slowing of replication fork progression, or both (Iyer and Rhind, 2013; Willis and Rhind, 2009b). Classically, the replication rate reduction in the setting of DNA damage has been thought to be directly modulated by the S-phase checkpoint response—that is, the checkpoint halts or delays replication to provide more time to repair DNA damage (Rhind and Russell, 2000). However, an alternative mechanism has been proposed, whereby the S-phase checkpoint actively triggers recombinational repair mechanisms first, which subsequently delay replication (Rhind and Russell, 2000). To investigate the mechanisms by which Cds1 mediates the reduction of bulk replication rate upon MMS treatment, our laboratory has focused on identifying Cds1 targets in fission yeast.

A limited number of Cds1 targets have been identified (Kai, 2003). Rad60, a recombinational repair protein, is phosphorylated by Cds1 upon hydroxyurea-induced replication fork arrest (Boddy et al., 2003; Miyabe et al., 2009). The *rad60-1* mutant strain is defective in double-strand break repair and is hypersensitive to MMS (Morishita et al., 2002). The Mus81-Eme1 complex in *S. pombe* is a structure-specific nuclease that resolves four-way DNA intermediates of homologous recombination known as Holliday junctions (Boddy et al., 2001). Cds1 phosphorylates Mus81 upon hydroxyurea treatment (Boddy et al., 2000). The *mus81.T239A* phosphomutant is unable to interact with Cds1(Kai, 2005). Rqh1 is a member of the RecQ family of DNA helicases. Rqh1 is a

homolog of the *S. cerevisiae* Sgs1 protein, and is related to the gene products encoded by human *BLM* and *WRN* that are mutated in Bloom's syndrome and Werner's syndrome respectively (Stewart et al., 1997). *rqh1*Δ cells treated with hydroxyurea have a cut (cell untimely torn) phenotype, which describes the appearance of a mitotic septum with sister chromosomes that are mis-segregated and unequally distributed due to aberrant mitosis (Stewart et al., 1997). The cut phenotype in hydroxyurea can result from cell cycle misregulation or checkpoint defects (Enoch and Nurse, 1990). Rqh1 appears to function in the S-phase checkpoint response and is important for preventing excessive recombination during hydroxyurea arrest (Chakraverty and Hickson, 1999; Murray et al., 1997). Expression of an *E. coli* Holliday junction resolvase in *rqh1*Δ fission yeast cells partially suppresses the UV and hydroxyurea hypersensitivities, and the cut phenotype of the *rqh1*Δ strain (Doe et al., 2000). Both the Mus81-Eme1 complex and the Rqh1 helicase appear to have roles in the repair of collapsed replication forks (Doe, 2002). Together, these Cds1 targets suggest that the S-phase checkpoint response to DNA damage involves recombination and repair.

Phenotypically, a number of mutants have been identified to have impaired S-phase checkpoint response. These mutants—in response to MMS treatment—do not properly reduce the rate of bulk DNA replication. *rad3*Δ and *cds1*Δ single mutants, which have a non-functional checkpoint sensor and effector kinases respectively, have defective slowing of replication upon MMS treatment. The Mre11-Rad50-Nbs1 (MRN) protein complex is involved in the sensing, processing and repair of DNA double-strand breaks (Rupnik et al., 2010; Williams et al., 2007). *rad32*Δ (Rad32 is a homolog of Mre11),

*rad50*Δ, and *nbs1*Δ single mutants exhibit a partial slowing of replication upon MMS treatment (Chahwan et al., 2003; Willis and Rhind, 2010).

Previous work in our laboratory sought to identify Cds1 targets responsible for MMS-induced slowing of replication. *mus81*Δ cells, *rqh1*Δ cells, *sfr1*Δ cells, and *rqh1*Δ *rhp55*Δ cells were reported to be defective in the slowing of replication in response to MMS. Sfr1 (Swi five-dependent recombination repair protein 1) is a component of the Swi5-Sfr1 protein complex that functions as an activator of DNA strand exchange in the Rad51-dependent homologous recombination pathway (Akamatsu et al., 2007, 2003; Argunhan et al., 2017; Haruta et al., 2006). Rhp55 is the homolog of *S. cerevisiae* Rad55, which is a component of the Rad55-Rad57 heterodimer that functions to stimulate Rad51-mediated strand exchange in the presence of replication protein A (RPA) (Khasanov et al., 1999; Sung, 1997a).

In *S. cerevisiae*, Rad52 and Rad51 are involved in the homologous recombination pathway of DNA double-strand break repair (Mortensen et al., 2009; Renkawitz et al., 2014). Rad52 mediates the loading of Rad51 on RPA-coated single-stranded DNA (Benson et al., 1998; New et al., 1998; Shinohara and Ogawa, 1998; Sung, 1997b). Rad51, a member of the RecA recombinase family, then performs homology search and promotes DNA strand invasion (Heyer et al., 2010; Kowalczykowski, 2015; Sung et al., 2000). Two homologs of Rad52 have been identified in *S. pombe*: Rad22A and Rti1 (also known as Rad22B). *rad22A*Δ and *rti1*Δ are synthetic lethal (Suto et al., 1999), but both the Rad22A and Rti1 proteins interact with RPA and Rhp51 (the homolog of *S.*

cerevisiae Rad51) and appear to have roles in recombination (Doe, 2004; Kim et al., 2002; Octobre et al., 2008; Tsutsui et al., 2001; van den Bosch et al., 2002, 2001).

To gain a better understanding of the S-phase checkpoint response, we sought to identify additional protein targets of Cds1 involved in the slowing of replication upon MMS treatment. A phosphoproteomics profiling study was conducted to identify phosphopeptides whose abundance increased in a Cds1-dependent manner upon MMS treatment (Willis et al., 2016). Briefly, this study arrested wildtype cells in G1 phase and released them into S-phase +/- MMS, harvested cells at mid-S phase, isolated peptides via trypsin digestion, differentially labelled peptides (light dimethyl label for MMS-treated, heavy dimethyl label for untreated control), and purified phosphopeptides for liquid chromatography combined with mass spectroscopy. The ratio of light-to-heavy labels provided the phosphopeptide enrichment in wildtype cells upon MMS treatment. The same approach was used to compare wildtype cells treated with MMS (light label) and *cds1*Δ cells treated with MMS (heavy label). From this, a list of Cds1-dependent, MMS-induced phosphopeptides was obtained.

A limitation of the phosphoproteomics profiling study is that the phosphopeptide enrichment (experimental over control samples) is not normalized to the total peptide abundance in each sample. The total peptide abundance is the sum of the phosphorylated and unphosphorylated peptides. As such, a 2-fold phosphopeptide enrichment in *cds1*Δ cells over wildtype cells—when both are treated with MMS—can result from at least two scenarios: the cells may produce the same amount of a particular peptide, but the peptide is phosphorylated with 2x frequency in *cds1*Δ cells; or the peptide is phosphorylated to

the same extent in both strains, but *cds1*Δ cells have 2x the total peptide expression levels. This limitation underscores the need to validate putative Cds1 targets.

To validate the putative Cds1 targets, we adopted a flow cytometry approach to assess the rate of bulk DNA replication. Mutants of candidate genes were synchronized, treated +/- MMS in S-phase, and samples at timed intervals were obtained to assess the DNA content. We hypothesized that null mutants of Cds1 targets would impair the checkpoint-mediated reduction of DNA replication rate upon MMS treatment. Thus, we screened mutants for their ability to recapitulate the *cds1*Δ checkpoint-deficient phenotype.

5.2. Materials and Methods

To develop a list of putative Cds1 targets to validate, the datasets from a phosphoproteomics profiling study were used (Willis et al., 2016). The previous study took wildtype cells and *cds1*Δ cells and identified phosphopeptides whose abundance were upregulated in S-phase upon MMS treatment. The relative phosphopeptide abundance enrichment of the wildtype strain over the *cds1*Δ strain was determined.

A list of putative Cds1 targets was generated based on the level of phosphopeptide abundance enrichment, the availability of the deletion strain, and the known function from literature and gene ontology (GO) annotations. Mutant strains from this list (**Table 5.1**) were obtained from the Bioneer (version 2) collection were crossed onto *cdc10-M17* background to facilitate G1 temperature-sensitive arrest (Kim et al., 2010). Cdc10 is a transcription factor required for the G1/S phase transition in S. pombe (Aves et al., 1985). The resulting strains were tested for MMS-induced S-phase slowing via bulk flow cytometry. Strains for *mus81*Δ, *rqh1*Δ, *sfr1*Δ, and *rqh1*Δ *rhp55*Δ were obtained from previous studies (Willis and Rhind, 2010, 2009a).

A flow cytometry based assay (Willis and Rhind, 2011) was used to assess the rate of bulk DNA replication. *S. pombe* cells were grown in liquid YES (yeast extract with supplements) at 25^{o}C and handled using standard methods (Forsburg and Rhind, 2006). At an optical density at 600nm (OD_{600}) of 1.0 +/- 0.15, mid-log phase growing cells were shifted to 35^{o}C for 2 hrs for *cdc10-M17* arrest. Afterwards, cells were loaded onto an elutriation chamber (preset at 35^{o}C) and the cells in the lowest density bracket

were collected at OD_{600} = 0.15 +/- 0.5. Collected cells were immediately divided into two flasks, treated +/- 0.03% MMS (Sigma-Aldrich), and maintained at 25^{o}C for replication. Cell aliquots were collected at 20 minute intervals, pelleted, and fixed in 70% ethanol for nuclei isolation. Isolation of nuclei was performed as described (Willis and Rhind, 2011) with two minor modifications. First, the amount of ethanol-fixed cells used at each timepoint for processing was 0.6 OD of cells. Second, the sonification step was performed using a Branson Sonifier 450 instrument with a microtip for 7 seconds at a constant duty cycle. Processed nuclei were stained with Sytox Green, and the DNA content was determined using flow cytometry.

The S-phase progression was calculated as described (Willis and Rhind, 2011) by determining how far the DNA content histogram peak at each timepoint has progressed from the 1C DNA content position to the 2C DNA content position. Specifically, at each timepoint, the S-phase progression was calculated using the formula: (C-A)/(B-A), where A = 1C DNA content position, B = 2C DNA content position, and C = histogram peak position at the particular timepoint. The 1C and 2C DNA content positions were determined from timepoints obtained prior to the S-phase release (i.e. from either the asynchronous sample and/or the 35^{o}C arrested sample. Thus, in a complete round of DNA replication, the S-phase progression increases from 0.0 (1C DNA content; unreplicated) to 1.0 (2C DNA content; fully replicated).

5.3. Results

A flow cytometry based assay (Willis and Rhind, 2011) was used to determine the S-phase checkpoint response in putative Cds1 targets. Cells synchronized in G1 phase and selected via centrifugal elutriation were allowed to replicate in either 0.03% MMS or no MMS. Samples were harvested over a 180-minute period and the DNA content was determined via flow cytometry.

Wildtype cells slow replication in MMS, whereas *cds1*Δ cells do not.

Wildtype (checkpoint-proficient) cells upon MMS treatment have reduced rate of bulk DNA replication over S-phase (i.e. they exhibit a replication-slowing phenotype). This reduction in the rate of bulk DNA replication was not present in *cds1*Δ (checkpoint-deficient) cells (**Figures 5.1 and 5.2**). Additionally, wildtype cells in MMS are unable to complete replication by t=180 minutes, whereas *cds1*Δ cells in MMS do complete replication. These results are consistent with previous findings that established Cds1 being the checkpoint effector kinase that facilitates the slowing of S-phase progression in the setting of MMS-induced DNA damage (Kumar and Huberman, 2004; Lindsay et al., 1998; Marchetti et al., 2002). The ability or inability of strains to slow bulk DNA replication in MMS thus serve as benchmarks for determining whether mutant strains are checkpoint-proficient or checkpoint-deficient respectively (**Table 5.1**).

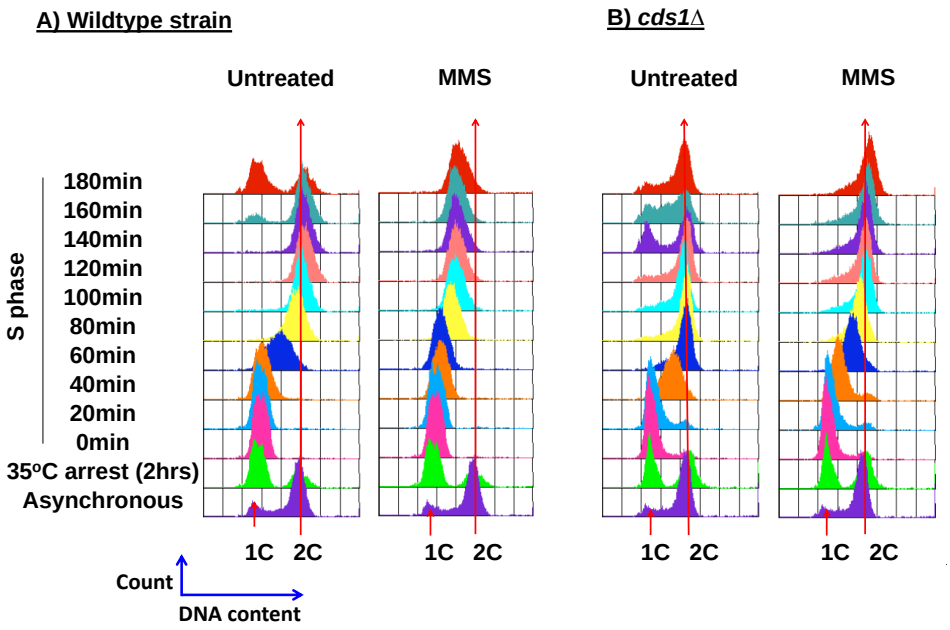


Figure 5.1. DNA content profile of synchronized (A) wildtype strain and (B) *cds1*Δ strain released into S-phase +/- 0.03% MMS.

The DNA content is determined by flow cytometry. Upon MMS treatment, wildtype (checkpoint-proficient) cells display a reduction in the rate of bulk replication (from 1C to 2C DNA content). Additionally, wildtype cells do not complete replication by t=180 minutes. In contrast, *cds1*Δ cells are checkpoint-deficient and complete replication (to 2C DNA content) in MMS.

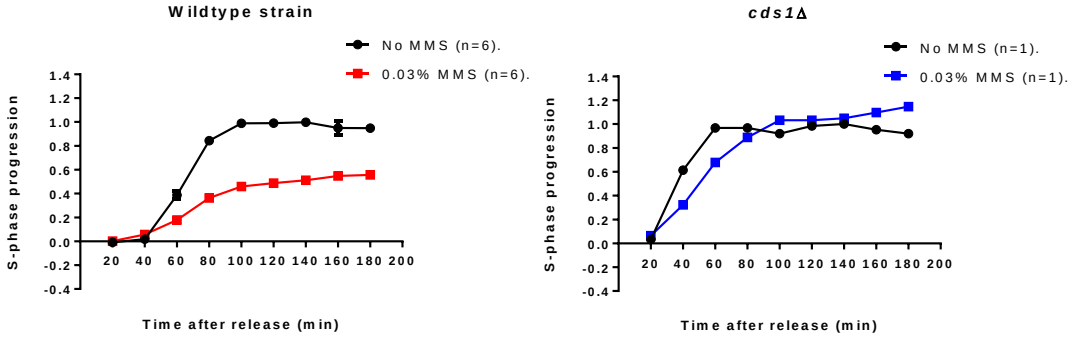


Figure 5.2. S-phase progression of wildtype strain and cds1Δ strain.

Wildtype cells treated with MMS slow bulk replication and do not complete replication by t=180 minutes. In contrast, *cds1*Δ cells complete replication. The S-phase progression indicates the position of the flow cytometry DNA content histogram peak as it migrates from the unreplicated/1C DNA content position (0.0) to the fully replicated/2C DNA content position (1.0). For the wildtype strain, the mean values are shown, with error bars indicating the standard error of the mean (SEM). In some cases, the error bars are not visible because the height of the error bar is shorter than the height of the symbol.

Table 5.1. Comparison of wildtype and cds1Δ phenotype upon MMS treatment.

Wildtype cells +MMS	***cds1*Δ cells +MMS**
Checkpoint proficient	Checkpoint deficient
Reduction in rate of bulk DNA replication	No reduction in rate of bulk DNA progression
Slowing of S-phase progression	No slowing of S-phase progression

To validate putative Cds1 phosphorylation targets, mutants of the candidate genes were treated +/- MMS, and their replication phenotypes were determined via flow cytometry. We began with the simplest model in which deletion of a *bona fide* Cds1 target would nullify the checkpoint response and recapitulate the *cds1*Δ phenotype (defective S-phase slowing). In this study, mostly single mutants were validated. However, since Cds1 may phosphorylate multiple targets, it is possible for single mutant phenotypes to be masked via compensatory mechanism or epistatic pathways.

All single mutants screened had reductions in the rate of bulk replication in MMS.

Fifteen single mutants from a list of putative Cds1 targets were evaluated for their ability to slow bulk replication upon MMS treatment (**Table 5.2**). Every mutant strain tested had a reduction in the rate of bulk DNA replication upon MMS treatment. This finding suggests that the strains are checkpoint-proficient. Among the mutant strains that had a reduction in bulk replication are *msh1*Δ, *msh3*Δ, *snf59*Δ, and *cdb4*Δ single mutants (**Figures 5.3 and 5.4**).

Table 5.2. List of strains that have reduced rate of bulk replication upon MMS treatment.

Gene product identity obtained from PomBase (McDowall et al., 2015; Wood et al., 2012).

Strain	Gene product
Wildtype	
Single mutants	
*rti1*Δ	A homolog of *S. cerevisiae* Rad52
*msh1*Δ	MutS protein homolog 1
*msh3*Δ	MutS protein homolog 3
*msh6*Δ	MutS protein homolog 6
*snf59*Δ	SWI/SNF complex subunit
*SPAC139.01c*Δ	Nuclease, XP-G family
*rhp7*Δ	Rad7 homolog
*rhp14*Δ	XP-A family homolog
*thp1*Δ	Uracil DNA N-glycosylase
*apn2*Δ	AP endonuclease 2
*rdh54*Δ	ATP-dependent DNA helicase
*mfh2*Δ	ATP-dependent DNA helicase
*exo1*Δ	Exonuclease I
*ctf18*Δ	Replication factor-C like complex subunit
*cdb4*Δ	Curved DNA binding protein
*SPBC28F2.11*Δ	HMG box
*sfr1*Δ	SWI five-dependent recombination mediator 1
*mus81*Δ	Holliday junction resolvase subunit
*rqh1*Δ	RecQ type DNA helicase
Double mutants	
*snf59*Δ *sfr1*Δ	Refer to above
*msh1*Δ *sfr1*Δ	Refer to above
*rti1*Δ *sfr1*Δ	Refer to above
*rqh1*Δ *rhp55*Δ	*Rhp55* is the homolog of S. cerevisiae Rad55

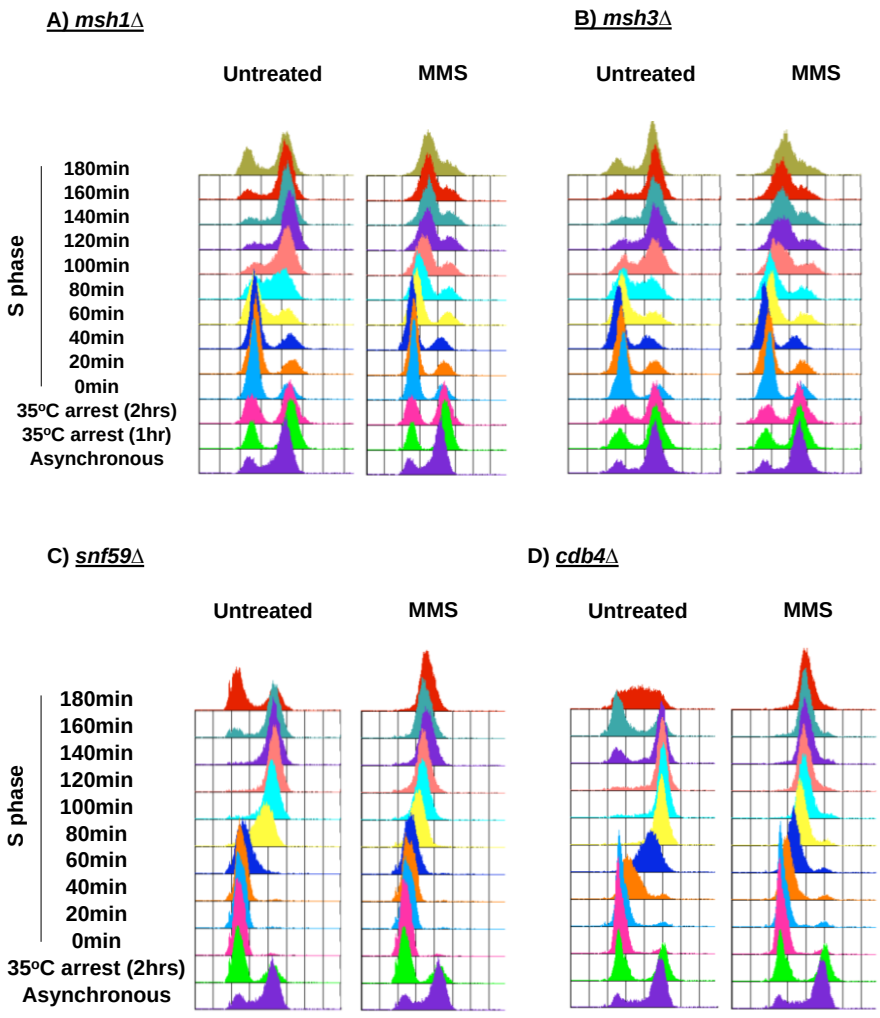


Figure 5.3. DNA content profile of six single mutant strains: *msh1*Δ; *msh3*Δ; *snf59*Δ; *cdb4*Δ; *rhp14*Δ; and *rdh54*Δ strains.

Upon MMS-treatment, all six strains display a reduction in the rate of bulk replication.

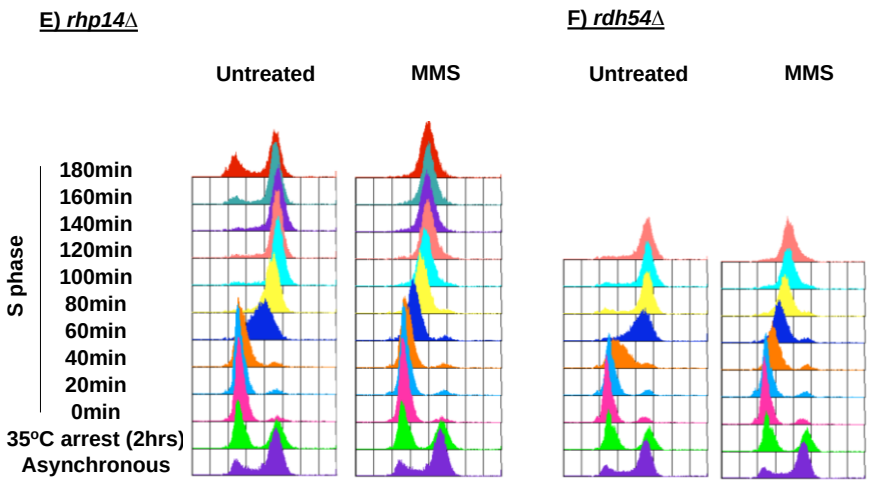


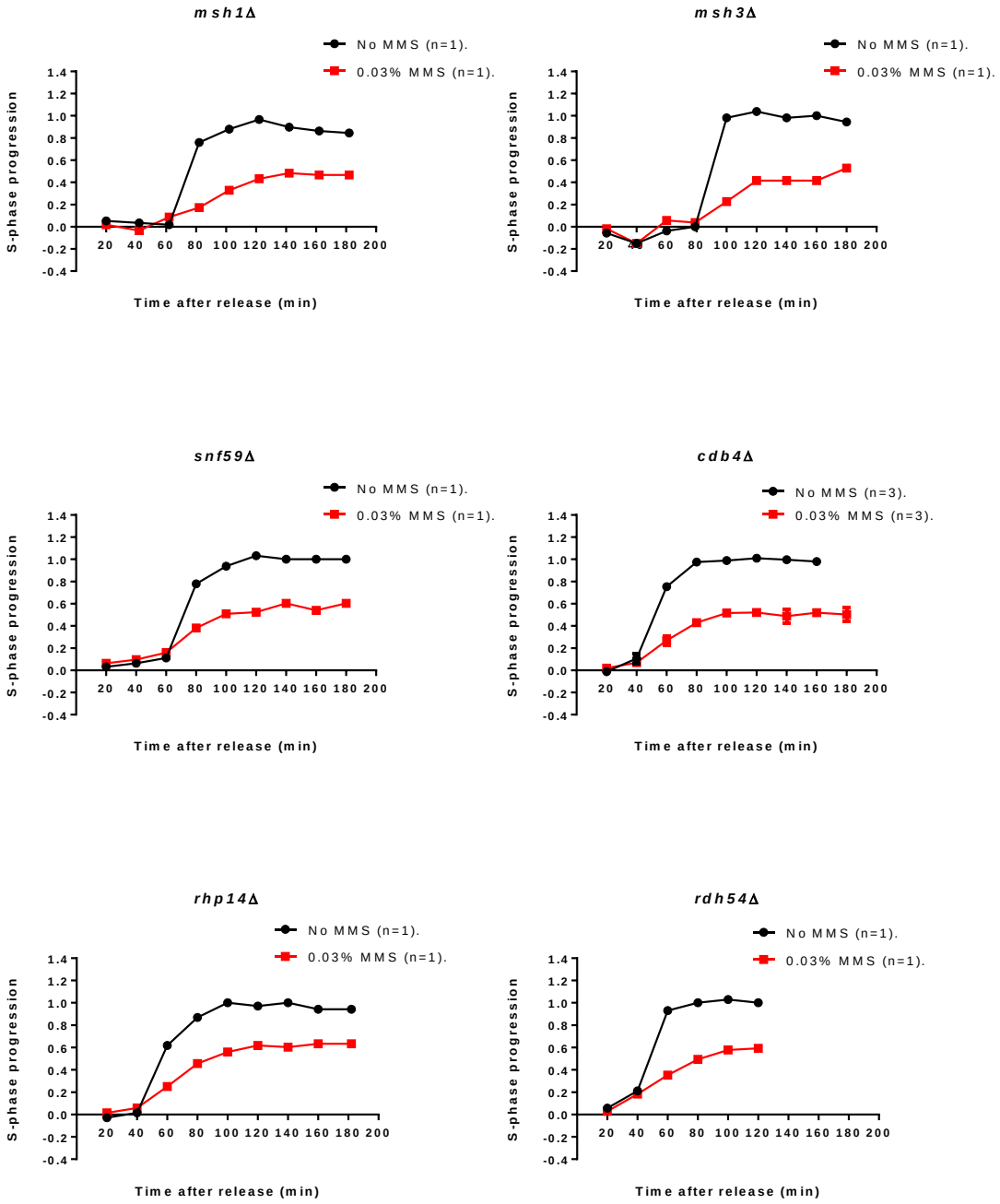
Figure 5.3 (continued). DNA content profile of six single mutant strains: A) *msh1*Δ; B) *msh3*Δ; C) *snf59*Δ; D) *cdb4*Δ; E) *rhp14*Δ; and F) *rdh54*Δ cells. Upon MMS-treatment, all six strains display a reduction in the rate of bulk replication.

Figure 5.4. S-phase progression of six single mutants: *msh1*Δ, *msh3*Δ, *snf59*Δ, *cdb4*Δ, *rhp14*Δ, and *rdh54*Δ strains.

Upon MMS treatment, all six strains slow bulk DNA replication and do not complete replication. For strains with replicates, the mean values are shown, with error bars indicating the standard error of the mean (SEM). In some cases, the error bars are not visible because the height of the error bar is shorter than the height of the symbol.

Figure on next page

Figure 5.4.


msh1Δ
No MMS (n=1).
0.03% MMS (n=1).
S-phase progression
Time after release (min)
msh3Δ
No MMS (n=1).
0.03% MMS (n=1).
S-phase progression
Time after release (min)
snf59Δ
No MMS (n=1).
0.03% MMS (n=1).
S-phase progression
Time after release (min)
cdb4Δ
No MMS (n=3).
0.03% MMS (n=3).
S-phase progression
Time after release (min)
rhp14Δ
No MMS (n=1).
0.03% MMS (n=1).
S-phase progression
Time after release (min)
rdh54Δ
No MMS (n=1).
0.03% MMS (n=1).
S-phase progression
Time after release (min)

Both *msh1* and *msh3* are predicted to encode for homologs of the *Escherichia coli* MutS protein that functions in DNA mismatch repair (McDowall et al., 2015; Wood et al., 2012). Snf59 is a component of the SWI/SNF chromatin remodeling complex (Monahan et al., 2008). Chromatin remodeling factors have been implicated in the S-phase checkpoint response in *S. cerevisiae*, as in the case of *isw2* and *nhp10* double mutants that slow replication upon MMS treatment (Au et al., 2011; Lee et al., 2015). *cdb4* encodes a protein that binds to curved DNA, but its function remains to be determined (Yamada et al., 1994). Rhp14 is a homolog of human XP-A and *S. cerevisiae* Rad14, which functions in nucleotide excision repair (Hohl et al., 2001). The *S. cerevisiae* Rdh54 interacts with Rad51 recombinase, promotes Rad51-mediated D-loop formation, and is involved in the removal of Rad51 (Chi et al., 2006, p. 54; Petukhova et al., 2000, 1998; Santa Maria et al., 2013; Shah et al., 2010). Additionally, the *S. pombe* Rdh54 and *S. cerevisiae* Rdh54 proteins both appear to have roles in recombinational repair of DNA double-strand breaks during meiosis (Catlett and Forsburg, 2003; Ferrari et al., 2013; Nimonkar et al., 2007; Pankratz and Forsburg, 2005; Tougan et al., 2010).

The effect of MMS-induced S-phase slowing is typically evident by 40min-80min after releasing cells into S-phase, as indicated by the delay in the flow cytometry peak migration from 1C DNA content (unreplicated) to 2C DNA content (fully replicated). We were thus unable to identify specific Cds1 targets using this bulk flow cytometry assay.

mus81, *rqh1*, and *sfr1* are not required for S-phase DNA damage checkpoint

The inability to identify any mutants that have a strong defect in the S-phase checkpoint response led us to investigate strains that were previously reported to be defective in replication slowing. *mus81*, *rqh1*, and *sfr1* were previously reported to be required for the S-phase checkpoint in response to MMS (Willis and Rhind, 2010, 2009a). Additionally, these two studies claimed that *mus81*Δ, *rqh1*Δ, *sfr1*Δ, and *rqh1*Δ *rhp55*Δ strains complete or almost complete replication (S-phase progression >0.9) by t=180 minutes in S-phase with MMS. Since those studies used the same bulk slowing assay, we sought to determine the reproducibility of their results.

Contrary to the aforementioned two studies, we found that *mus81*Δ, *rqh1*Δ, and *sfr1*Δ single mutants—which were obtained from those two studies—all exhibited S-phase slowing in response to MMS. Additionally, the strains do not complete replication by t=180 minutes in S-phase with MMS. Their phenotypes were much closer to wildtype than the *cds1*Δ phenotype, and thus demonstrated checkpoint proficiency (**Figures 5.5 and 5.6**). The *rqh1*Δ result here is supported by another study, which demonstrated that *rqh1*Δ exhibited S-phase slowing in response to MMS, akin to the wildtype strain (Marchetti et al., 2002). These results are in contrast to two reports (Willis and Rhind, 2010, 2009a), which concluded that MMS-induced S-phase slowing is defective in *rqh1*Δ, *sfr1*Δ and *mus81*Δ mutants.

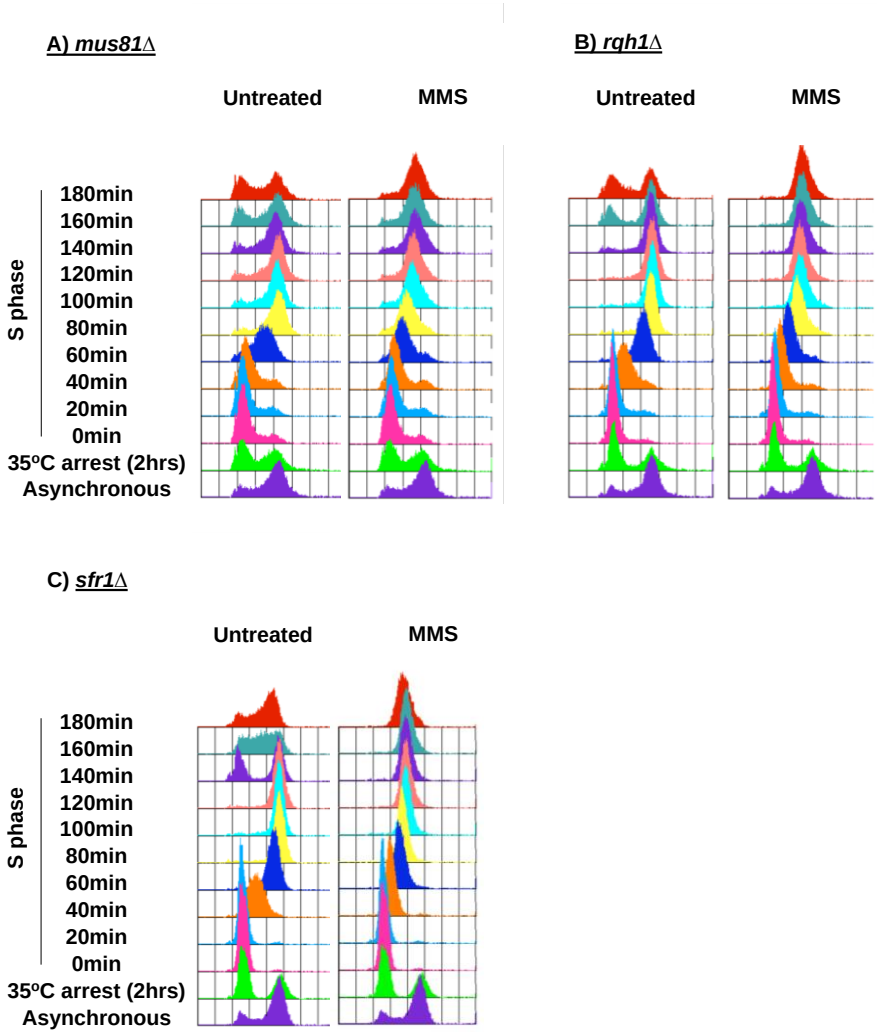


Figure 5.5. DNA content profiles of three strains: A) *mus81*Δ; B) *rqh1*Δ; and C) *sfr1*Δ strains.

Upon MMS-treatment, all three strains display a reduction in the rate of bulk replication.

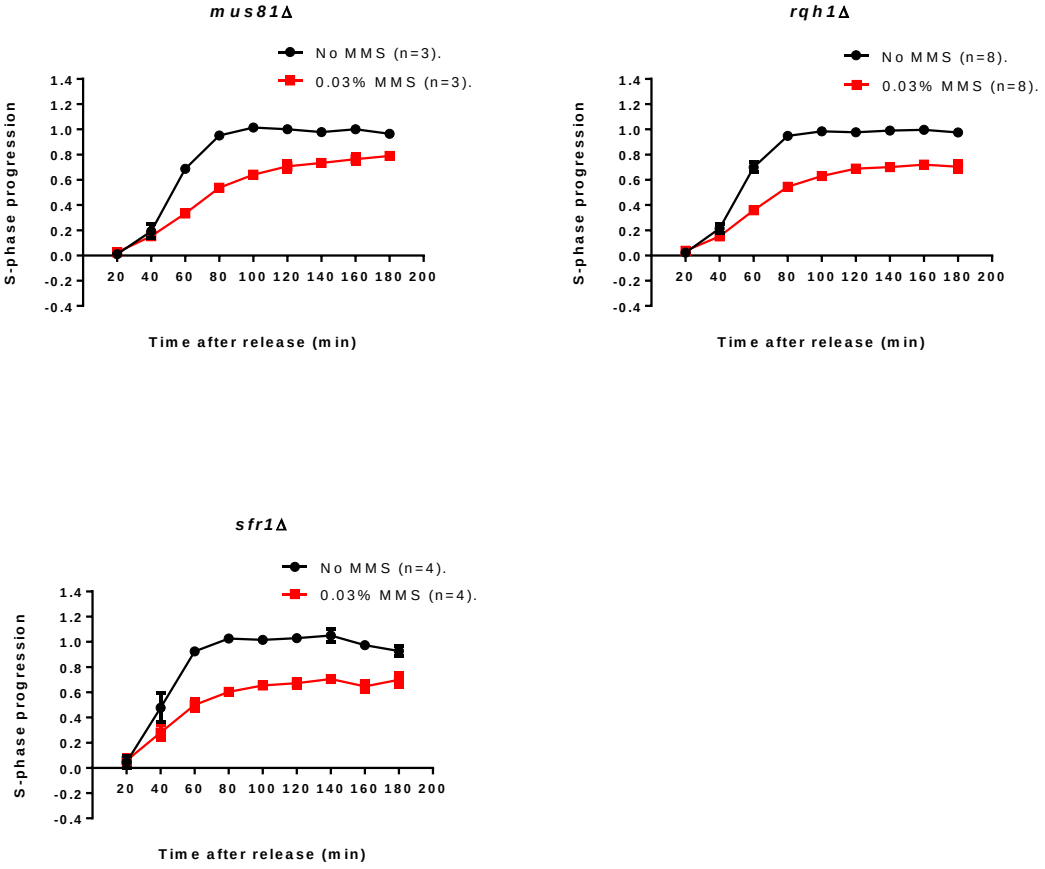


Figure 5.6. S-phase progression of *mus81*Δ, *rqh1*Δ, and *sfr1*Δ strains.

Upon MMS treatment, all these strains slow bulk DNA replication and do not complete replication. For strains with replicates, the mean values are shown, with error bars indicating the standard error of the mean (SEM). In some cases, the error bars are not visible because the height of the error bar is shorter than the height of the symbol.

Four double mutants have a reduction of bulk replication rate in MMS

Four double mutants examined in this study exhibited S-phase slowing in response to MMS, and thus demonstrated checkpoint proficiency: *snf59*Δ *sfr1*Δ, *msh1*Δ *sfr1*Δ, *rti1*Δ *sfr1*Δ, and *rqh1*Δ *rhp55*Δ (**Figures 5.7 and 5.8**). We found that *rqh1*Δ *rhp55*Δ did not complete replication in MMS (S-phase progression = 0.65), contrary to a previous study that claimed it did (S-phase progression > 0.9) (Willis and Rhind, 2009a).

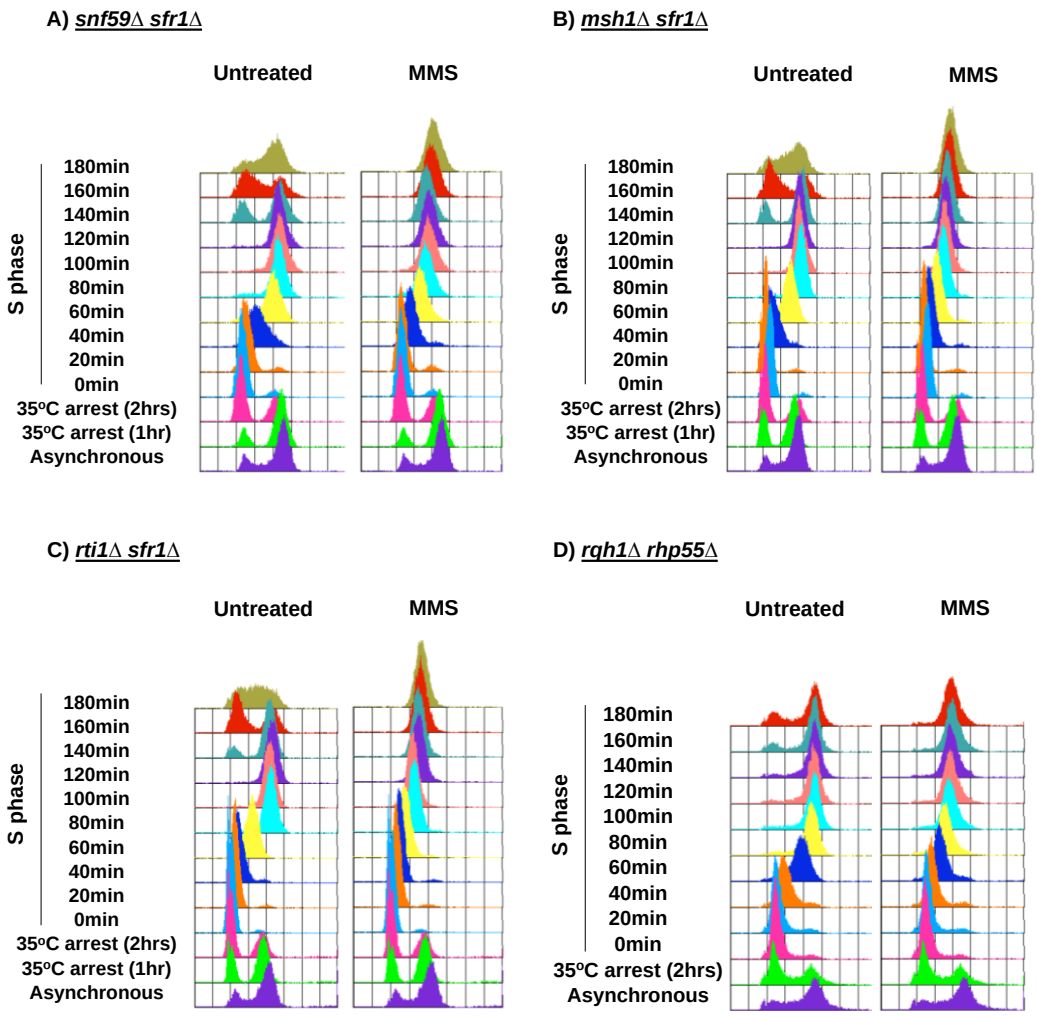


Figure 5.7. DNA content profiles of four double mutant strains: A) *snf59*Δ *sfr1*Δ; B) *msh1*Δ *sfr1*Δ; C) *rti1*Δ *sfr1*Δ; and D) *rqh1*Δ *rhp55*Δ strains.

Upon MMS-treatment, all four strains display a reduction in the rate of bulk replication.

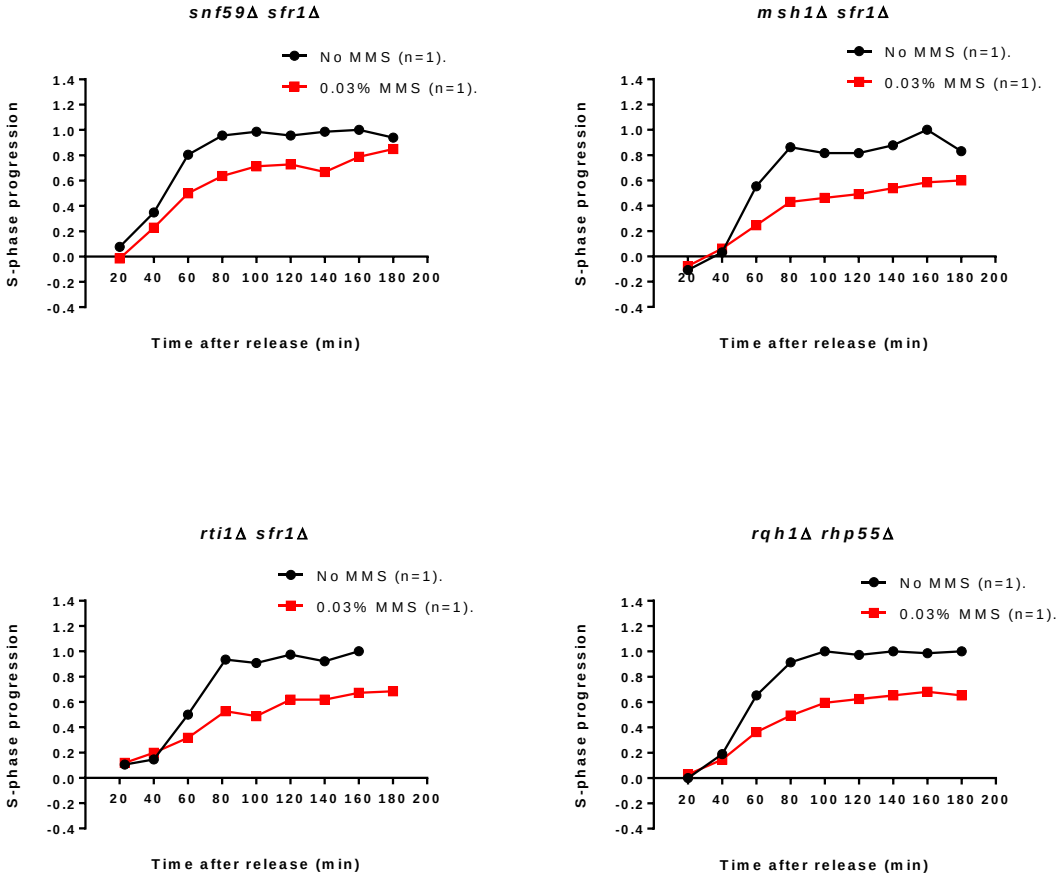


Figure 5.8. S-phase progression of double mutants: *snf59*Δ *sfr1*Δ; *msh1*Δ *sfr1*Δ; *rti1*Δ *sfr1*Δ; and *rqh1*Δ *rhp55*Δ strains.

Upon MMS treatment, all these strains slow bulk DNA replication.

Comparison of strains

A comparison of the bulk replication rate for wildtype, *mus81*Δ, *rqh1*Δ, *sfr1*Δ, and *rqh1*Δ *rhp55*Δ strains is shown in **Figure 5.9**. The S-phase progression of the strains at t=140min in S-phase, Rep_{140}, is shown in **Figure 5.10**. With the exception of the *cds1*Δ control, which is checkpoint deficient, none of the strains investigated here completed replication by t=180minutes in MMS.

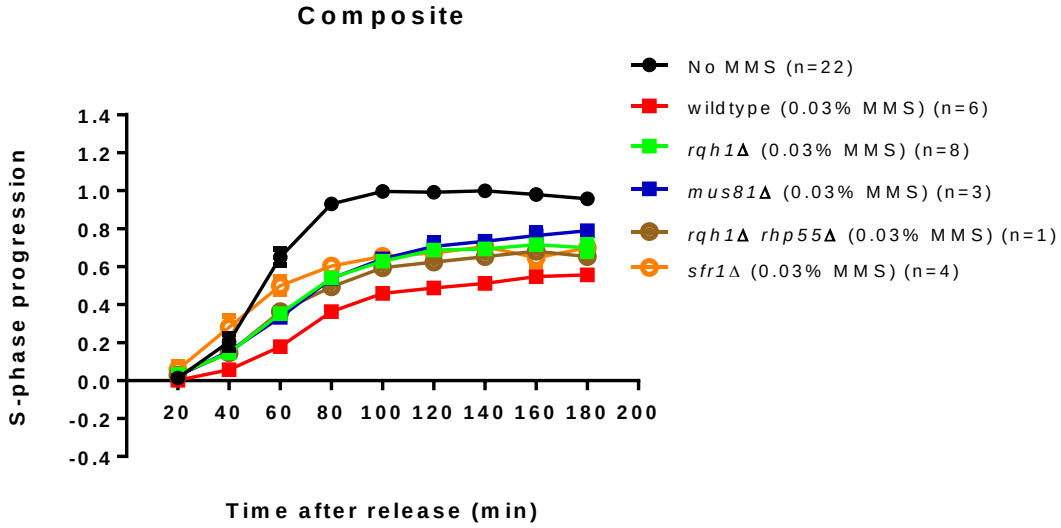


Figure 5.9. S-phase progression composite of wildtype; *rqh1*Δ; *mus81*Δ; *rqh1*Δ *rhp55*Δ; and *sfr1*Δ strains.

Upon MMS treatment, all these strains slow bulk DNA replication. For strains with replicates, the mean values are shown, with error bars indicating the standard error of the mean (SEM). In some cases, the error bars are not visible because the height of the error bar is shorter than the height of the symbol.

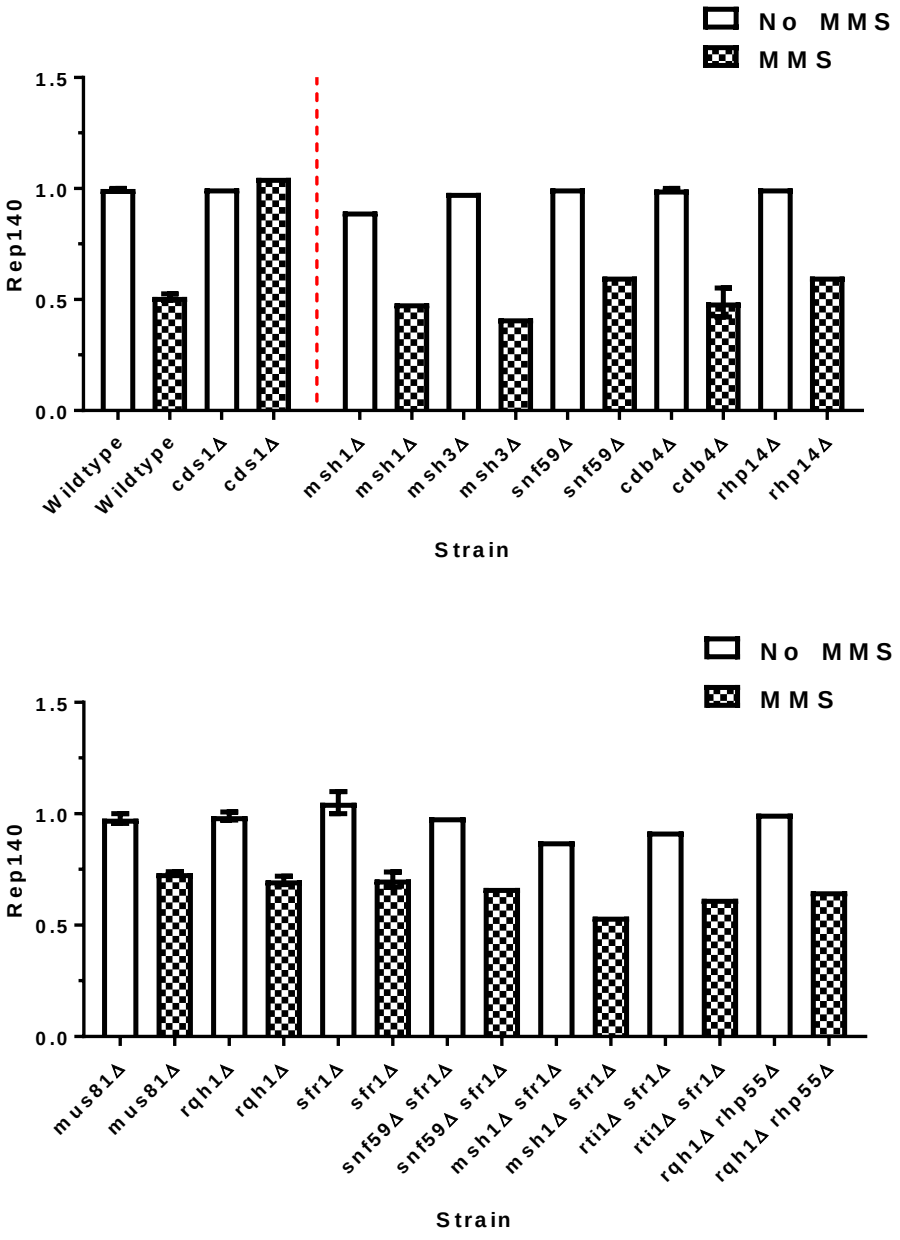


Figure 5.10. Summary of replication data showing Rep_{140}.

Rep_{140} is the S-phase progression value at t=140 minutes after S-phase release. For strains where 3 or more experiments were performed, the mean values are shown, with error bars indicating the standard error of the mean (SEM).

5.4. Discussion

In this study, we sought to identify protein targets of Cds1, the S-phase checkpoint effector kinase in *S. pombe,* in the setting of MMS-induced DNA damage. A flow cytometry assay was used to monitor S-phase progression of mutant strains, from 1C DNA content (unreplicated) to 2C DNA content (fully replicated). Consistent with previous findings, wildtype (checkpoint-proficient) cells treated with MMS displayed a slowing of bulk DNA replication (Kumar and Huberman, 2004; Marchetti et al., 2002). In contrast, *cds1*Δ (checkpoint-deficient) cells treated with MMS did not slow replication, as expected (Kumar and Huberman, 2004; Marchetti et al., 2002).

Of the 15 candidate gene mutants examined, all displayed a slowing of S-phase progression phenotype upon MMS treatment. Thus, based on the single mutant studies, we were unable to identify Cds1 targets using the bulk flow cytometry assay. However, since there may be multiple Cds1 targets, epistatic pathways may mask the change in phenotype of null alleles of individual *bona fide* Cds1 targets. A more sophisticated investigation, involving double mutants and multiple mutants, may yield strains that recapitulate the *cds1*Δ phenotype.

Our inability to identify any novel Cds1 targets led us to repeat the assay in *rqh1*Δ, *sfr1*Δ, *mus81*Δ, and *rqh1*Δ *rhp55*Δ mutants—strains that were previously reported to be defective in their ability to reduce the bulk replication rate in MMS. However, my conclusion is at variance with the two previous studies (Willis and Rhind, 2010, 2009a). In those studies, the authors claimed that *rqh1*Δ, *sfr1*Δ, *mus81*Δ, and *rqh1*Δ *rhp55*Δ

strains are checkpoint defective—that is, they do not slow replication upon MMS treatment. Of note is that the two studies did not report any original flow cytometry profiles for these strains to support their claim. Instead of reporting the raw data, the studies only reported the calculated S-phase progression. My finding that *rqh1*Δ does have reduced replication rate upon MMS is in agreement with another study (Marchetti et al., 2002).

Together, the findings here indicate that none of the single mutants screened recapitulate the bulk replication phenotype of the checkpoint-deficient *rad3*Δ or *cds1*Δ strains. This study suggests a more complex pathway in which multiple Cds1 targets may work in concert to slow the progression of DNA replication in response to MMS-induced DNA damage. The development of high-throughput approaches to study replication kinetics on single DNA molecules may facilitate the identification of Cds1 targets and provide a greater understanding of the S-phase checkpoint.

Chapter VI: Concluding remarks

6.1. Concluding remarks

DNA replication is a key biological process. The replicon model of DNA replication (Jacob et al., 1963) can explain the origin behavior in *E. coli* and classically in *S. cerevisiae*. However, our understanding of origin behavior in higher eukaryotes remains limited. This is partly due to the complexity of higher eukaryotic origins, including cell-to-cell variability in origin usage and the low efficiency of many origins. Existing single molecule approaches to studying DNA replication—such as molecular combing with FISH—are low throughput and call for novel approaches to study replication origins and replication kinetics.

Here, I explored the feasibility of using PacBio SMRT-sequencing and the BioNano Genomics Irys optical mapping platform to detect sites of replication. Additionally, I discussed the prospect of using nanopore sequencing to study DNA replication. As the evidence here suggests, these three approaches are highly promising and may advance our understanding of DNA replication. A number of experiments can be performed to accelerate the development of these innovative approaches.

In Chapter II, I demonstrated the potential of using SMRT sequencing to detect thymidine analogs, which can be used for pulse labeling cells to mark sites of active replication. There, I showed that the presence BrdU, EdU, or IdU in a synthetic template can be detected via analysis of the kinetics of base incorporation (**Figure 2.7**). The power of this approach is exemplified by the ability to distinguish a single EdU from thymidine (**Figure 2.8**). Given its sensitivity, it is worthy to also explore the ability to detect a single

thymidine-to-BrdU and a single thymidine-to-IdU substitution. BrdU and IdU are cell-tolerable nucleosides commonly used in replication studies.

Another cell-tolerable nucleoside analog that should be explored is AmdU. Although not described in Chapter II, I have successfully incorporated AmdU into synthetic DNA templates and performed SMRT sequencing. Moreover, I have shown that the AmdU can be conjugated to BCN-OH or DBCO-OH in a post-synthetic reaction, yielding DNA templates of lower gel electrophoretic mobility (**Figure 3.15**). AmdU conjugated to a bulky group may conceivably affect base incorporation kinetics further, and thus enhance the signal for detection. Other nucleoside analogs, especially non-thymidine analogs, should be also be explored.

SMRT-sequencing can also be performed using pulse labeled genomic DNA. In synchronized cells where the pulse labeling is restricted to one S-phase, SMRTbell template preparations would yield two DNA template strands: one strand with labels, and one strand unlabeled. To determine sites of active replication, SMRTbells with high IPD values in the labeled strand can be identified. To identify the genomic source of SMRTbells with high IPD values, the base-called DNA sequence from the unlabeled strand can be used. With optimization of the PacBio base-caller algorithm, it may even be possible to accurately determine the genomic source from the labeled strand. Moreoever, it may be possible to identify pulse labels at a single-base resolution. This can increase the accuracy and precision in identifying replication origins and in calculating replication fork speed.

In Chapter III, I demonstrated the potential of using the BioNano Genomics Irys platform to optically map replication origins. The analysis pipeline continues to be optimized for replication studies, but it would be promising to conduct pulse labeling experiments with AmdU. Since AmdU-containing DNA can be conjugated to BCN- or DBCO-containing compounds (**Figure 3.15**), it is highly plausible that BCN- or DBCO-containing fluorophores can be detected. Conceivably, if multiple colors can be detected in nanofluidic instruments, all three labeling strategies presented in Chapter III can be combined to map origins.

Although not described in this thesis, we have also worked with Nabsys in developing approaches to study replication on single DNA molecules. As new single molecule mapping technologies from Nabsys, Genia Technologies, and others emerge, it is highly plausible that the work developed in this thesis can facilitate the rapid growth of applying these technologies to study DNA replication. For example, the work that led to the successful preparation of synthetic DNA can now be quickly repeated to test the feasibility of new technology platforms for studying DNA replication. Similarly, the *in vivo* pulse labeling experiments—especially with AmdU +/- conjugation—can be performed to quickly provide pulse labeled DNA for sequencing.

Single molecule approaches continue to be developed for the study of replication origins and replication kinetics. As these approaches progress up the technology S-curve (Sood, 2010), it would be fascinating to see how our understanding of DNA replication rapidly evolves in our lifetime, which is only a tiny fraction of the human race.

References

Abdurashidova, G., Deganuto, M., Klima, R., Riva, S., Biamonti, G., Giacca, M., Falaschi, A., 2000. Start sites of bidirectional DNA synthesis at the human lamin B2 origin. Science 287, 2023–2026.

Agard, N.J., Prescher, J.A., Bertozzi, C.R., 2004. A Strain-Promoted [3 + 2] Azide−Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. J. Am. Chem. Soc. 126, 15046–15047.

Akamatsu, Y., Dziadkowiec, D., Ikeguchi, M., Shinagawa, H., Iwasaki, H., 2003. Two different Swi5-containing protein complexes are involved in mating-type switching and recombination repair in fission yeast. Proc. Natl. Acad. Sci. 100, 15770–15775.

Akamatsu, Y., Tsutsui, Y., Morishita, T., Siddique, M.S.P., Kurokawa, Y., Ikeguchi, M., Yamao, F., Arcangioli, B., Iwasaki, H., 2007. Fission yeast Swi5/Sfr1 and Rhp55/Rhp57 differentially regulate Rhp51-dependent recombination outcomes. EMBO J. 26, 1352–1362.

Aladjem, M.I., Rodewald, L.W., Kolman, J.L., Wahl, G.M., 1998. Genetic dissection of a mammalian replicator in the human beta-globin locus. Science 281, 1005–1009.

Allen, J.B., Zhou, Z., Siede, W., Friedberg, E.C., Elledge, S.J., 1994. The SAD1/RAD53 protein kinase controls multiple checkpoints and DNA damage-induced transcription in yeast. Genes Dev. 8, 2401–2415.

Argunhan, B., Murayama, Y., Iwasaki, H., 2017. The differentiated and conserved roles of Swi5-Sfr1 in homologous recombination. FEBS Lett. 591, 2035–2047.

Aruoma, O.I., Halliwell, B., Gajewski, E., Dizdaroglu, M., 1991. Copper-ion-dependent damage to the bases in DNA in the presence of hydrogen peroxide. Biochem. J. 273, 601–604.

Au, T.J., Rodriguez, J., Vincent, J.A., Tsukiyama, T., 2011. ATP-Dependent Chromatin Remodeling Factors Tune S Phase Checkpoint Activity. Mol. Cell. Biol. 31, 4454–4463.

Austin, R.J., Orr-Weaver, T.L., Bell, S.P., 1999. Drosophila ORC specifically binds to ACE3, an origin of DNA replication control element. Genes Dev. 13, 2639–2649.

Avery, O.T., MacLeod, C.M., McCarty, M., 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J. Exp. Med. 79, 137.

Aves, S.J., Durkacz, B.W., Carr, A., Nurse, P., 1985. Cloning, sequencing and transcriptional control of the Schizosaccharomyces pombe cdc10'start'gene. EMBO J. 4, 457.

Azmi, I.F., Watanabe, S., Maloney, M.F., Kang, S., Belsky, J.A., MacAlpine, D.M., Peterson, C.L., Bell, S.P., 2017. Nucleosomes influence multiple steps during replication initiation. eLife 6, e22512.

Bartholdy, B., Mukhopadhyay, R., Lajugie, J., Aladjem, M.I., Bouhassira, E.E., 2015. Allele-specific analysis of DNA replication origins in mammalian cells. Nat. Commun. 6, 7051.

Baskin, J.M., Prescher, J.A., Laughlin, S.T., Agard, N.J., Chang, P.V., Miller, I.A., Lo, A., Codelli, J.A., Bertozzi, C.R., 2007. Copper-free click chemistry for dynamic in vivo imaging. Proc. Natl. Acad. Sci. 104, 16793–16797.

Bayley, H., 2015. Nanopore Sequencing: From Imagination to Reality. Clin. Chem. 61, 25–31.

Beach, D., Piper, M., Shall, S., 1980. Isolation of chromosomal origins of replication in yeast. Nature 284, 185–187.

Bechhoefer, J., Rhind, N., 2012. Replication timing and its emergence from stochastic processes. Trends Genet. 28, 374–381.

Bell, S.P., 2002. The origin recognition complex: from simple origins to complex functions. Genes Dev. 16, 659–672.

Bell, S.P., 2016. Mechanisms of Chromosomal DNA Replication: Initiation of DNA Replication. iBiology.org. Available at: www.ibiology.org/ibioseminars/mechanisms-chromosomal-dna-replication-initiation-dna-replication.html (accessed May 15, 2017).

Bell, S.P., Dutta, A., 2002. DNA Replication in Eukaryotic Cells. Annu. Rev. Biochem. 71, 333–374.

Bell, S.P., Stillman, B., 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature 357, 128–134.

Belsky, J.A., MacAlpine, H.K., Lubelsky, Y., Hartemink, A.J., MacAlpine, D.M., 2015. Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly. Genes Dev. 29, 212–224.

Bensimon, A., Simon, A., Chiffaudel, A., Croquette, V., Heslot, F., Bensimon, D., 1994. Alignment and sensitive detection of DNA by a moving interface. Science 265, 2096–2098.

Benson, F.E., Baumann, P., West, S.C., 1998. Synergistic actions of Rad51 and Rad52 in recombination and DNA repair. Nature 391, 401–404.

Bentley, N.J., Holtzman, D.A., Flaggs, G., Keegan, K.S., DeMaggio, A., Ford, J.C., Hoekstra, M., Carr, A.M., 1996. The Schizosaccharomyces pombe rad3 checkpoint gene. EMBO J. 15, 6641.

Beranek, D.T., 1990. Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. Mutat. Res. 231, 11–30.

Berbenetz, N.M., Nislow, C., Brown, G.W., 2010. Diversity of Eukaryotic DNA Replication Origins Revealed by Genome-Wide Analysis of Chromatin Structure. PLoS Genet. 6, e1001092.

Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J.-M., Lemaitre, J.-M., 2012. Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nat. Struct. Mol. Biol. 19, 837–844.

Bevilacqua, V., King, M., Chaumontet, M., Nothisen, M., Gabillet, S., Buisson, D., Puente, C., Wagner, A., Taran, F., 2014. Copper-Chelating Azides for Efficient Click Conjugation Reactions in Complex Media. Angew. Chem. Int. Ed. 53, 5872–5876.

Biamonti, G., Giacca, M., Perini, G., Contreas, G., Zentilin, L., Weighardt, F., Guerra, M., Della Valle, G., Saccone, S., Riva, S., 1992. The gene for a novel human lamin maps at a highly transcribed locus of chromosome 19 which replicates at the onset of S-phase. Mol. Cell. Biol. 12, 3499–3506.

Bianco, J.N., Poli, J., Saksouk, J., Bacal, J., Silva, M.J., Yoshida, K., Lin, Y.-L., Tourrière, H., Lengronne, A., Pasero, P., 2012. Analysis of DNA replication profiles in budding yeast and mammalian cells using DNA combing. Methods 57, 149–157.

Bielinsky, A.K., Gerbi, S.A., 1998. Discrete start sites for DNA synthesis in the yeast ARS1 origin. Science 279, 95–98.

Bird, R.E., Louarn, J., Martuscelli, J., Caro, L., 1972. Origin and sequence of chromosome replication in Escherichia coli. J. Mol. Biol. 70, 549–566.

Blattner, F.R., Plunkett, G. 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997. The complete genome sequence of Escherichia coli K-12. Science 277, 1453–1462.

Blow, J.J., Gillespie, P.J., Francis, D., Jackson, D.A., 2001. Replication Origins in XenopusEgg Extract Are 5–15 Kilobases Apart and Are Activated in Clusters That Fire at Different Times. J. Cell Biol. 152, 15–26.

Boddy, M.N., Gaillard, P.-H.L., McDonald, W.H., Shanahan, P., Yates, J.R., Russell, P., 2001. Mus81-Eme1 are essential components of a Holliday junction resolvase. Cell 107, 537–548.

Boddy, M.N., Lopez-Girona, A., Shanahan, P., Interthal, H., Heyer, W.-D., Russell, P., 2000. Damage Tolerance Protein Mus81 Associates with the FHA1 Domain of Checkpoint Kinase Cds1. Mol. Cell. Biol. 20, 8758–8766.

Boddy, M.N., Shanahan, P., McDonald, W.H., Lopez-Girona, A., Noguchi, E., Yates III, J.R., Russell, P., 2003. Replication Checkpoint Kinase Cds1 Regulates Recombinational Repair Protein Rad60. Mol. Cell. Biol. 23, 5939–5946.

Bonhoeffer, F., Gierer, A., 1963. On the growth mechanism of the bacterial chromosome. J. Mol. Biol. 7, 534–540.

Bramhill, D., Kornberg, A., 1988. Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the E. coli chromosome. Cell 52, 743–755.

Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S.B., Krstic, P.S., Lindsay, S., Ling, X.S., Mastrangelo, C.H., Meller, A., Oliver, J.S., Pershin, Y.V., Ramsey, J.M., Riehn, R., Soni, G.V., Tabard-Cossa, V., Wanunu, M., Wiggin, M., Schloss, J.A., 2008. The potential and challenges of nanopore sequencing. Nat. Biotechnol. 26, 1146–1153.

Brewer, B.J., Fangman, W.L., 1987. The localization of replication origins on ARS plasmids in S. cerevisiae. Cell 51, 463–471.

Brown, C.G., Clarke, J., 2016. Nanopore development at Oxford Nanopore. Nat Biotech 34, 810–811.

Buettner, G.R., Jurkiewicz, B.A., 1996. Catalytic metals, ascorbate and free radicals: combinations to avoid. Radiat. Res. 145, 532–541.

Buschmann, T., Zhang, R., Brash, D.E., Bystrykh, L.V., 2014. Enhancing the detection of barcoded reads in high throughput DNA sequencing data by controlling the false discovery rate. BMC Bioinformatics 15, 264.

Bußkamp, H., Batroff, E., Niederwieser, A., Abdel-Rahman, O.S., Winter, R.F., Wittmann, V., Marx, A., 2014. Efficient labelling of enzymatically synthesized vinyl-modified DNA by an inverse-electron-demand Diels–Alder reaction. Chem Commun 50, 10827–10829.

Butler, T.Z., Pavlenok, M., Derrington, I.M., Niederweis, M., Gundlach, J.H., 2008. Single-molecule DNA detection with an engineered MspA protein nanopore. Proc. Natl. Acad. Sci. 105, 20647–20652.

Caddie, M.S., Calos, M.P., 1992. Analysis of the autonomous replication behavior in human cells of the dihydrofolate reductase putative chromosomal origin of replication. Nucleic Acids Res. 20, 5971–5978.

Cairns, J., 1963. The bacterial chromosome and its manner of replication as seen by autoradiography. J. Mol. Biol. 6, 208IN3–213IN5.

Cao, Hongzhi, Hastie, A.R., Cao, D., Lam, E.T., Sun, Y., Huang, H., Liu, X., Lin, L., Andrews, W., Chan, S., Huang, S., Tong, X., Requa, M., Anantharaman, T., Krogh, A., Yang, H., Cao, Han, Xu, X., 2014. Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. GigaScience 3, 34.

Cappella, P., Giansanti, V., Pulici, M., Gasparri, F., 2013. From “Click” to “Fenton” chemistry for 5-bromo-2′-deoxyuridine determination: Fenton Chemistry for BrdU Incorporation Detection. Cytometry A 83, 989–1000.

Carminati, J.L., Johnston, C.G., Orr-Weaver, T.L., 1992. The Drosophila ACE3 chorion element autonomously induces amplification. Mol. Cell. Biol. 12, 2444–2453.

Carter, J.-M., Hussain, S., 2017. Robust long-read native DNA sequencing using the ONT CsgG Nanopore system. Wellcome Open Res. 2.

Catlett, M.G., Forsburg, S.L., 2003. Schizosaccharomyces pombe Rdh54 (TID1) acts with Rhp54 (RAD54) to repair meiotic double-strand breaks. Mol. Biol. Cell 14, 4707–4720.

Cayrou, C., Ballester, B., Peiffer, I., Fenouil, R., Coulombe, P., Andrau, J.-C., van Helden, J., Méchali, M., 2015. The chromatin environment shapes DNA replication origin organization and defines origin classes. Genome Res. 25, 1873–1885.

Cayrou, C., Coulombe, P., Puy, A., Rialle, S., Kaplan, N., Segal, E., Méchali, M., 2012. New insights into replication origin characteristics in metazoans. Cell Cycle 11, 658–667.

Chahwan, C., Nakamura, T.M., Sivakumar, S., Russell, P., Rhind, N., 2003. The Fission Yeast Rad32 (Mre11)-Rad50-Nbs1 Complex Is Required for the S-Phase DNA Damage Checkpoint. Mol. Cell. Biol. 23, 6564–6573.

Chakraborty, T., Yoshinaga, K., Lother, H., Messer, W., 1982. Purification of the E. coli dnaA gene product. EMBO J. 1, 1545.

Chakraverty, R.K., Hickson, I.D., 1999. Defending genome integrity during DNA replication: a proposed role for RecQ family helicases. BioEssays News Rev. Mol. Cell. Dev. Biol. 21, 286–294.

Chan, C.S., Tye, B.-K., 1980. Autonomously replicating sequences in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. 77, 6329–6333.

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

Chi, P., Kwon, Y., Seong, C., Epshtein, A., Lam, I., Sung, P., Klein, H.L., 2006. Yeast Recombination Factor Rdh54 Functionally Interacts with the Rad51 Recombinase and Catalyzes Rad51 Removal from DNA. J. Biol. Chem. 281, 26268–26279.

Chuang, R.-Y., Kelly, T.J., 1999. The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. Proc. Natl. Acad. Sci. 96, 2656–2661.

Cimprich, K.A., Shin, T.B., Keith, C.T., Schreiber, S.L., 1996. cDNA cloning and gene mapping of a candidate human cell cycle checkpoint protein. Proc. Natl. Acad. Sci. 93, 2850–2855.

Clark, T.A., Spittle, K.E., Turner, S.W., Korlach, J., 2011. Direct detection and sequencing of damaged DNA bases. Genome Integr. 2, 10.

Clarke, J., Wu, H.-C., Jayasinghe, L., Patel, A., Reid, S., Bayley, H., 2009. Continuous base identification for single-molecule nanopore DNA sequencing. Nat. Nanotechnol. 4, 265–270.

Claycomb, J.M., MacAlpine, D.M., Evans, J.G., Bell, S.P., Orr-Weaver, T.L., 2002. Visualization of replication initiation and elongation in *Drosophila*. J. Cell Biol. 159, 225–236.

Conti, C., Saccà, B., Herrick, J., Lalou, C., Pommier, Y., Bensimon, A., 2007. Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells. Mol. Biol. Cell 18, 3059–3067.

Cross, J.B., Currier, R.P., Torraco, D.J., Vanderberg, L.A., Wagner, G.L., Gladen, P.D., 2003. Killing of Bacillus Spores by Aqueous Dissolved Oxygen, Ascorbic Acid, and Copper Ions. Appl. Environ. Microbiol. 69, 2245–2252.

Czajkowsky, D.M., Liu, J., Hamlin, J.L., Shao, Z., 2008. DNA Combing Reveals Intrinsic Temporal Disorder in the Replication of Yeast Chromosome VI. J. Mol. Biol. 375, 12–19.

Dahmann, C., Diffley, J.F., Nasmyth, K.A., 1995. S-phase-promoting cyclin-dependent kinases prevent re-replication by inhibiting the transition of replication origins to a pre-replicative state. Curr. Biol. 5, 1257–1269.

Dai, J., Chuang, R.-Y., Kelly, T.J., 2005. DNA replication origins in the Schizosaccharomyces pombe genome. Proc. Natl. Acad. Sci. U. S. A. 102, 337–342.

Das, S.K., Austin, M.D., Akana, M.C., Deshpande, P., Cao, H., Xiao, M., 2010. Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. Nucleic Acids Res. 38, e177–e177.

De Carli, F., Gaggioli, V., Millot, G.A., Hyrien, O., 2016. Single-molecule, antibody-free fluorescent visualisation of replication tracts along barcoded DNA molecules. Int. J. Dev. Biol.

De Castro, E., Soriano, I., Marín, L., Serrano, R., Quintales, L., Antequera, F., 2012. Nucleosomal organization of replication origins and meiotic recombination hotspots in fission yeast. EMBO J. 31, 124–137.

Deamer, D., Akeson, M., Branton, D., 2016. Three decades of nanopore sequencing. Nat. Biotechnol. 34, 518–524.

Delidakis, C., Kafatos, F.C., 1989. Amplification enhancers and replication origins in the autosomal chorion gene cluster of Drosophila. EMBO J. 8, 891.

Dellino, G.I., Cittaro, D., Piccioni, R., Luzi, L., Banfi, S., Segalla, S., Cesaroni, M., Mendoza-Maldonado, R., Giacca, M., Pelicci, P.G., 2013. Genome-wide mapping of human DNA-replication origins: Levels of transcription at ORC1 sites regulate origin selection and replication timing. Genome Res. 23, 1–11.

DePamphilis, M.L., 1999. Replication origins in metazoan chromosomes: fact or fiction? BioEssays News Rev. Mol. Cell. Dev. Biol. 21, 5–16.

Derrington, I.M., Butler, T.Z., Collins, M.D., Manrao, E., Pavlenok, M., Niederweis, M., Gundlach, J.H., 2010. Nanopore DNA sequencing with MspA. Proc. Natl. Acad. Sci. 107, 16060–16065.

Deshpande, A.M., Newlon, C.S., 1992. The ARS consensus sequence is required for chromosomal origin function in Saccharomyces cerevisiae. Mol. Cell. Biol. 12, 4305–4313.

Devbhandari, S., Jiang, J., Kumar, C., Whitehouse, I., Remus, D., 2017. Chromatin Constrains the Initiation and Elongation of DNA Replication. Mol. Cell 65, 131–141.

Diffley, J.F., Stillman, B., 1988. Purification of a yeast protein that binds to origins of DNA replication and a transcriptional silencer. Proc. Natl. Acad. Sci. 85, 2120–2124.

Ding, Y., Fleming, A.M., White, H.S., Burrows, C.J., 2015. Differentiation of G:C *vs* A:T and G:C *vs* G:mC Base Pairs in the Latch Zone of α-Hemolysin. ACS Nano 9, 11325–11332.

Doe, C.L., 2002. Mus81-Eme1 and Rqh1 Involvement in Processing Stalled and Collapsed Replication Forks. J. Biol. Chem. 277, 32753–32759.

Doe, C.L., 2004. DNA repair by a Rad22-Mus81-dependent pathway that is independent of Rhp51. Nucleic Acids Res. 32, 5570–5581.

Doe, C.L., Dixon, J., Osman, F., Whitby, M.C., 2000. Partial suppression of the fission yeast rqh1- phenotype by expression of a bacterial Holliday junction resolvase. EMBO J. 19, 2751–2762.

Drury, L.S., Perkins, G., Diffley, J.F., 2000. The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. Curr. Biol. 10, 231–240.

Dubey, D.D., Davis, L.R., Greenfeder, S.A., Ong, L.Y., Zhu, J.G., Broach, J.R., Newlon, C.S., Huberman, J.A., 1991. Evidence suggesting that the ARS elements associated with silencers of the yeast mating-type locus HML do not function as chromosomal DNA replication origins. Mol. Cell. Biol. 11, 5346–5355.

Duzdevich, D., Warner, M.D., Ticau, S., Ivica, N.A., Bell, S.P., Greene, E.C., 2015. The Dynamics of Eukaryotic Replication Initiation: Origin Specificity, Licensing, and Firing at the Single-Molecule Level. Mol. Cell 58, 483–494.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., MacAlpine, D.M., 2010. Conserved nucleosome positioning defines replication origins. Genes Dev. 24, 748–753.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J., Turner, S., 2009. Real-time DNA sequencing from single polymerase molecules. Science 323, 133–138.

Eki, T., Murakami, Y., Hanaoka, F., 2013. Trapping DNA Replication Origins from the Human Genome. Genes 4, 198–225.

Elford, H.L., 1968. Effect of hydroxyurea on ribonucleotide reductase. Biochem. Biophys. Res. Commun. 33, 129–135.

Enoch, T., Nurse, P., 1990. Mutation of fission yeast cell cycle control genes abolishes dependence of mitosis on DNA replication. Cell 60, 665–673.

Eshaghi, M., Karuturi, R.K.M., Li, J., Chu, Z., Liu, E.T., Liu, J., 2007. Global Profiling of DNA Replication Timing and Efficiency Reveals that Efficient Replication/Firing Occurs Late during S-Phase in S. pombe. PLoS ONE 2, e722.

Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B., Speck, C., 2009. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. Proc. Natl. Acad. Sci. 106, 20240–20245.

Faller, M., Niederweis, M., Schulz, G.E., 2004. The Structure of a Mycobacterial Outer-Membrane Channel. Science 303, 1189.

Fang, G., Munera, D., Friedman, D.I., Mandlik, A., Chao, M.C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M.C., Jabado, O.J., Deikus, G., Clark, T.A., Luong, K., Murray, I.A., Davis, B.M., Keren-Paz, A., Chess, A., Roberts, R.J., Korlach, J., Turner, S.W., Kumar, V., Waldor, M.K., Schadt, E.E., 2012. Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing. Nat. Biotechnol. 30, 1232–1239.

Feng, Y.-Q., Warin, R., Li, T., Olivier, E., Besse, A., Lobell, A., Fu, H., Lin, C.M., Aladjem, M.I., Bouhassira, E.E., 2005. The Human -Globin Locus Control Region Can Silence as Well as Activate Gene Expression. Mol. Cell. Biol. 25, 3864–3874.

Ferrari, M., Nachimuthu, B.T., Donnianni, R.A., Klein, H., Pellicioli, A., 2013. Tid1/Rdh54 translocase is phosphorylated through a Mec1- and Rad53-dependent manner in the presence of DSB lesions in budding yeast. DNA Repair 12, 347–355.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., Turner, S.W., 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat. Methods 7, 461–465.

Forsburg, S.L., Rhind, N., 2006. Basic methods for fission yeast. Yeast 23, 173–183.

Foulk, M.S., Urban, J.M., Casella, C., Gerbi, S.A., 2015. Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. Genome Res. 25, 725–735.

Fragkos, M., Ganier, O., Coulombe, P., Méchali, M., 2015. DNA replication origin activation in space and time. Nat. Rev. Mol. Cell Biol. 16, 360–374.

Fuller, C.W., Kumar, S., Porel, M., Chien, M., Bibillo, A., Stranges, P.B., Dorwart, M., Tao, C., Li, Z., Guo, W., Shi, S., Korenblum, D., Trans, A., Aguirre, A., Liu, E., Harada, E.T., Pollard, J., Bhat, A., Cech, C., Yang, A., Arnold, C., Palla, M., Hovis, J., Chen, R., Morozova, I., Kalachikov, S., Russo, J.J., Kasianowicz, J.J., Davis, R., Roever, S., Church, G.M., Ju, J., 2016. Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array. Proc. Natl. Acad. Sci. 113, 5233–5238.

Fuller, R.S., Funnell, B.E., Kornberg, A., 1984. The dnaA protein complex with the E. coli chromosomal replication origin (oriC) and other DNA sites. Cell 38, 889–900.

Genest, P.-A., Baugh, L., Taipale, A., Zhao, W., Jan, S., van Luenen, H.G.A.M., Korlach, J., Clark, T., Luong, K., Boitano, M., Turner, S., Myler, P.J., Borst, P., 2015. Defining the sequence requirements for the positioning of base J in DNA using SMRT sequencing. Nucleic Acids Res. 43, 2102–2115.

Gerbi, S.A., Bielinsky, A.-K., 1997. Replication initiation point mapping. Methods 13, 271–280.

Giacca, M., Pelizon, C., Falaschi, A., 1997. Mapping replication origins by quantifying relative abundance of nascent DNA strands using competitive polymerase chain reaction. Methods San Diego Calif 13, 301–312.

Giacca, M., Zentilin, L., Norio, P., DIvIACCO, S., Dimitrova, D., Contreas, G., Biamonti, G., Perini, G., Weighardt, F., Riva, S., 1994. Fine mapping of a replication origin of human DNA. Proc. Natl. Acad. Sci. 91, 7119–7123.

Gilbert, D.M., 2010. Evaluating genome-scale approaches to eukaryotic DNA replication. Nat. Rev. Genet. 11, 673–684.

Givens, R.M., Lai, W.K.M., Rizzo, J.M., Bard, J.E., Mieczkowski, P.A., Leatherwood, J., Huberman, J.A., Buck, M.J., 2012. Chromatin architectures at fission yeast transcriptional promoters and replication origins. Nucleic Acids Res. 40, 7176–7189.

Gouaux, E., 1998. α-Hemolysin fromStaphylococcus aureus: an archetype of β-barrel, channel-forming toxins. J. Struct. Biol. 121, 110–122.

Goyal, P., Krasteva, P.V., Van Gerven, N., Gubellini, F., Van den Broeck, I., Troupiotis-Tsaïlaki, A., Jonckheere, W., Péhau-Arnaudet, G., Pinkner, J.S., Chapman, M.R., Hultgren, S.J., Howorka, S., Fronzes, R., Remaut, H., 2014. Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. Nature 516, 250–253.

Graham, J.E., Marians, K.J., Kowalczykowski, S.C., 2017. Independent and Stochastic Action of DNA Polymerases in the Replisome. Cell 169, 1201–1213.e17.

Gratzner, H., 1982. Monoclonal antibody to 5-bromo- and 5-iododeoxyuridine: A new reagent for detection of DNA replication. Science 218, 474.

Hamlin, J.L., Mesner, L.D., Dijkwel, P.A., 2010. A winding road to origin discovery. Chromosome Res. 18, 45–61.

Harland, R.M., Laskey, R.A., 1980. Regulated replication of DNA microinjected into eggs of Xenopus laevis. Cell 21, 761–771.

Haruta, N., Kurokawa, Y., Murayama, Y., Akamatsu, Y., Unzai, S., Tsutsui, Y., Iwasaki, H., 2006. The Swi5–Sfr1 complex stimulates Rhp51/Rad51 - and Dmc1-mediated DNA strand exchange in vitro. Nat. Struct. Mol. Biol. 13, 823–830.

Hastie, A.R., Dong, L., Smith, A., Finklestein, J., Lam, E.T., Huo, N., Cao, H., Kwok, P.-Y., Deal, K.R., Dvorak, J., Luo, M.-C., Gu, Y., Xiao, M., 2013. Rapid Genome Mapping in Nanochannel Arrays for Highly Complete and Accurate De Novo Sequence Assembly of the Complex Aegilops tauschii Genome. PLoS ONE 8, e55864.

Hayashi, M., Katou, Y., Itoh, T., Tazumi, M., Yamada, Y., Takahashi, T., Nakagawa, T., Shirahige, K., Masukata, H., 2007. Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. EMBO J. 26, 1327–1339.

Heck, M.M., Spradling, A.C., 1990. Multiple replication origins are used during Drosophila chorion gene amplification. J. Cell Biol. 110, 903–914.

Heichinger, C., Penkett, C.J., Bähler, J., Nurse, P., 2006. Genome-wide characterization of fission yeast DNA replication origins. EMBO J. 25, 5171.

Heinzel, S.S., Krysan, P.J., Tran, C.T., Calos, M.P., 1991. Autonomous DNA replication in human cells is affected by the size and the source of the DNA. Mol. Cell. Biol. 11, 2263–2272.

Henke, W., Herdel, K., Jung, K., Schnorr, D., Loening, S.A., 1997. Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Res. 25, 3957–3958.

Herrick, J., Stanislawski, P., Hyrien, O., Bensimon, A., 2000. Replication fork density increases during DNA synthesis in X. laevis egg extracts11Edited by M. Yaniv. J. Mol. Biol. 300, 1133–1142.

Hershey, A.D., Chase, M., 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J. Gen. Physiol. 36, 39–56.

Heyer, W.-D., Ehmsen, K.T., Liu, J., 2010. Regulation of Homologous Recombination in Eukaryotes. Annu. Rev. Genet. 44, 113–139.

Hoeijmakers, J.H., 2009. DNA damage, aging, and cancer. N. Engl. J. Med. 361, 1475–1485.

Hohl, M., Christensen, O., Kunz, C., Naegeli, H., Fleck, O., 2001. Binding and Repair of Mismatched DNA Mediated by Rhp14, the Fission Yeast Homologue of Human XPA. J. Biol. Chem. 276, 30766–30772.

Huberman, J.A., Riggs, A.D., 1968. On the mechanism of DNA replication in mammalian chromosomes. J. Mol. Biol. 32, 327IN15335–334IN22341.

Huberman, J.A., Spotila, L.D., Nawotka, K.A., El-Assouli, S.M., Davis, L.R., 1987. The in vivo replication origin of the yeast 2μm plasmid. Cell 51, 473–481.

Hublin, J.-J., Ben-Ncer, A., Bailey, S.E., Freidline, S.E., Neubauer, S., Skinner, M.M., Bergmann, I., Le Cabec, A., Benazzi, S., Harvati, K., Gunz, P., 2017. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. Nature 546, 289–292.

Hyrien, O., Marheineke, K., Goldar, A., 2003. Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. BioEssays 25, 116–125.

Hyrien, O., Mechali, M., 1993. Chromosomal replication initiates and terminates at random sequences but at regular intervals in the ribosomal DNA of Xenopus early embryos. EMBO J. 12, 4511.

Imlay, J.A., Linn, S., 1988. DNA damage and oxygen radical toxicity. Science 240, 1302–1309.

Iyer, D.R., Das, S., Rhind, N., 2017. Analysis of DNA Replication in Fission Yeast by Combing. Cold Spring Harb. Protoc.

Iyer, D.R., Rhind, N., 2013. Checkpoint regulation of replication forks: global or local? Biochem. Soc. Trans. 41, 1701–1705.

Jackson, D.A., Pombo, A., 1998. 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. 140, 1285–1295.

Jacob, F., Brenner, S., Cuzin, F., 1963. On the Regulation of DNA Replication in Bacteria. Cold Spring Harb. Symp. Quant. Biol. 28, 329–348.

Jacob, F., Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318–356.

Jensen, M.A., Fukushima, M., Davis, R.W., 2010. DMSO and Betaine Greatly Improve Amplification of GC-Rich Constructs in De Novo Synthesis. PLoS ONE 5, e11024.

Jiang, H., Zheng, T., Lopez-Aguilar, A., Feng, L., Kopp, F., Marlow, F.L., Wu, P., 2014. Monitoring Dynamic Glycosylation in Vivo Using Supersensitive Click Chemistry. Bioconjug. Chem. 25, 698–706.

Jo, K., Dhingra, D.M., Odijk, T., de Pablo, J.J., Graham, M.D., Runnheim, R., Forrest, D., Schwartz, D.C., 2007. A single-molecule barcoding system using nanoslits for DNA analysis. Proc. Natl. Acad. Sci. 104, 2673–2678.

Jun, S., Rhind, N., 2008. Just-in-time DNA replication. Physics 1.

Kaguni, J.M., 2006. DnaA: Controlling the Initiation of Bacterial DNA Replication and More. Annu. Rev. Microbiol. 60, 351–371.

Kai, M., 2003. Checkpoint responses to replication stalling: inducing tolerance and preventing mutagenesis. Mutat. Res. Mol. Mech. Mutagen. 532, 59–73.

Kai, M., 2005. Replication checkpoint kinase Cds1 regulates Mus81 to preserve genome integrity during replication stress. Genes Dev. 19, 919–932.

Kasianowicz, J.J., Brandin, E., Branton, D., Deamer, D.W., 1996. Characterization of individual polynucleotide molecules using a membrane channel. Proc. Natl. Acad. Sci. 93, 13770–13773.

Keith, C.T., Schreiber, S.L., 1995. PIK-related kinases: DNA repair, recombination, and cell cycle checkpoints. Science 270, 50.

Kelly, T.J., Stillman, B., 2006. Duplication of DNA in eukaryotic cells. In: DNA Replication and Human Disease, Cold Spring Harbor Monograph Series. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 1–30.

Khasanov, F.K., Savchenko, G.V., Bashkirova, E.V., Korolev, V.G., Heyer, W.-D., Bashkirov, V.I., 1999. A new recombinational DNA repair gene from Schizosaccharomyces pombe with homology to Escherichia coli RecA. Genetics 152, 1557–1572.

Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., Duhig, T., Nam, M., Palmer, G., Han, S., Jeffery, L., Baek, S.-T., Lee, H., Shim, Y.S., Lee, M., Kim, L., Heo, K.-S., Noh, E.J., Lee, A.-R., Jang, Y.-J., Chung, K.-S., Choi, S.-J., Park, J.-Y., Park, Y., Kim, H.M., Park, S.-K., Park, H.-J., Kang, E.-J., Kim, H.B., Kang, H.-S., Park, H.-M., Kim, K., Song, K., Song, K.B., Nurse, P., Hoe, K.-L., 2010. Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. 28, 617–623.

Kim, W.J., Park, E.J., Lee, H., Seong, R.H., Park, S.D., 2002. Physical Interaction between Recombinational Proteins Rhp51 and Rad22 in *Schizosaccharomyces pombe*. J. Biol. Chem. 277, 30264–30270.

Kitsberg, D., Selig, S., Keshet, I., Cedar, H., 1993. Replication structure of the human beta-globin gene domain. Nature 366, 588–590.

Kolb, H.C., Finn, M.G., Sharpless, K.B., 2001. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. Angew. Chem. Int. Ed Engl. 40, 2004–2021.

Korlach, J., Bjornson, K.P., Chaudhuri, B.P., Cicero, R.L., Flusberg, B.A., Gray, J.J., Holden, D., Saxena, R., Wegener, J., Turner, S.W., 2010. Real-Time DNA Sequencing from Single Polymerase Molecules. In: Methods in Enzymology. Elsevier, pp. 431–455.

Korlach, J., Turner, S.W., 2012. Going beyond five bases in DNA sequencing. Curr. Opin. Struct. Biol. 22, 251–261.

Kowalczykowski, S.C., 2015. An Overview of the Molecular Mechanisms of Recombinational DNA Repair. Cold Spring Harb. Perspect. Biol. 7, a016410.

Krysan, P.J., Calos, M.P., 1991. Replication initiates at multiple locations on an autonomously replicating plasmid in human cells. Mol. Cell. Biol. 11, 1464–1472.

Kumar, S., Huberman, J.A., 2004. On the Slowing of S Phase in Response to DNA Damage in Fission Yeast. J. Biol. Chem. 279, 43574–43580.

Kumar, S., Tao, C., Chien, M., Hellner, B., Balijepalli, A., Robertson, J.W.F., Li, Z., Russo, J.J., Reiner, J.E., Kasianowicz, J.J., Ju, J., 2012. PEG-Labeled Nucleotides and Nanopore Detection for Single Molecule DNASequencing by Synthesis. Sci. Rep. 2.

Kurat, C.F., Yeeles, J.T.P., Patel, H., Early, A., Diffley, J.F.X., 2017. Chromatin Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates. Mol. Cell 65, 117–130.

Lam, E.T., Hastie, A., Lin, C., Ehrlich, D., Das, S.K., Austin, M.D., Deshpande, P., Cao, H., Nagarajan, N., Xiao, M., Kwok, P.-Y., 2012. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat. Biotechnol. 30, 771–776.

Lantermann, A.B., Straub, T., Strålfors, A., Yuan, G.-C., Ekwall, K., Korber, P., 2010. Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. Nat. Struct. Mol. Biol. 17, 251–257.

Larkin, J., Foquet, M., Turner, S.W., Korlach, J., Wanunu, M., 2014. Reversible Positioning of Single Molecules inside Zero-Mode Waveguides. Nano Lett. 14, 6023–6029.

Laskey, R.A., 1985. Chromosome replication in early development of Xenopus laevis. Development 89, 285–296.

Laszlo, A.H., Derrington, I.M., Brinkerhoff, H., Langford, K.W., Nova, I.C., Samson, J.M., Bartlett, J.J., Pavlenok, M., Gundlach, J.H., 2013. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. Proc. Natl. Acad. Sci. 110, 18904–18909.

Lee, L., Rodriguez, J., Tsukiyama, T., 2015. Chromatin Remodeling Factors Isw2 and Ino80 Regulate Checkpoint Activity and Chromatin Structure in S Phase. Genetics 199, 1077–1091.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., Nislow, C., 2007. A high-resolution atlas of nucleosome occupancy in yeast. Nat. Genet. 39, 1235–1244.

Lengronne, A., Pasero, P., Bensimon, A., Schwob, E., 2001. Monitoring S phase progression globally and locally using BrdU incorporation in TK+ yeast strains. Nucleic Acids Res. 29, 1433–1442.

Liang, C., Stillman, B., 1997. Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in cdc6 mutants. Genes Dev. 11, 3375–3386.

Lindsay, H.D., Griffiths, D.J., Edwards, R.J., Christensen, P.U., Murray, J.M., Osman, F., Walworth, N., Carr, A.M., 1998. S-phase-specific activation of Cds1 kinase defines a subpathway of the checkpoint response in Schizosaccharomyces pombe. Genes Dev. 12, 382–395.

Lipford, J.R., Bell, S.P., 2001. Nucleosomes positioned by ORC facilitate the initiation of DNA replication. Mol. Cell 7, 21–30.

Liu, D.S., Tangpeerachaikul, A., Selvaraj, R., Taylor, M.T., Fox, J.M., Ting, A.Y., 2012. Diels–Alder Cycloaddition for Fluorophore Targeting to Specific Proteins inside Living Cells. J. Am. Chem. Soc. 134, 792–795.

Liu, P.-Y., Jiang, N., Zhang, J., Wei, X., Lin, H.-H., Yu, X.-Q., 2006. The oxidative damage of plasmid DNA by ascorbic acid derivatives in vitro: the first research on the relationship between the structure of ascorbic acid and the oxidative damage of plasmid DNA. Chem. Biodivers. 3, 958–966.

Liu, S., Wall, T.A., Ozcelik, D., Parks, J.W., Hawkins, A.R., Schmidt, H., 2015. Electro-optical detection of single λ-DNA. Chem Commun 51, 2084–2087.

Liu, S., Zhao, Y., Parks, J.W., Deamer, D.W., Hawkins, A.R., Schmidt, H., 2014. Correlated Electrical and Optical Analysis of Single Nanoparticles and Biomolecules on a Nanopore-Gated Optofluidic Chip. Nano Lett. 14, 4816–4820.

Lu, H., Giordano, F., Ning, Z., 2016. Oxford Nanopore MinION Sequencing and Genome Assembly. Genomics Proteomics Bioinformatics 14, 265–279.

Lu, L., Zhang, H., Tower, J., 2001. Functionally distinct, sequence-specific replicator and origin elements are required for Drosophila chorion gene amplification. Genes Dev. 15, 134–146.

Lubelsky, Y., Sasaki, T., Kuipers, M.A., Lucas, I., Le Beau, M.M., Carignon, S., Debatisse, M., Prinz, J.A., Dennis, J.H., Gilbert, D.M., 2011. Pre-replication complex proteins assemble at regions of low nucleosome occupancy within the Chinese hamster dihydrofolate reductase initiation zone. Nucleic Acids Res. 39, 3141–3155.

Lucas, I., Chevrier-Miller, M., Sogo, J.M., Hyrien, O., 2000. Mechanisms ensuring rapid and complete DNA replication despite random initiation in Xenopus early embryos. J. Mol. Biol. 296, 769–786.

Luzzietti, N., Knappe, S., Richter, I., Seidel, R., 2012. Nicking enzyme–based internal labeling of DNA at multiple loci. Nat. Protoc. 7, 643–653.

Maaløe, O., Hanawalt, P.C., 1961. Thymine deficiency and the normal DNA replication cycle. I. J. Mol. Biol. 3, 144–155.

MacAlpine, H.K., Gordân, R., Powell, S.K., Hartemink, A.J., MacAlpine, D.M., 2010. Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. Genome Res. 20, 201–211.

Mak, A.C.Y., Lai, Y.Y.Y., Lam, E.T., Kwok, T.-P., Leung, A.K.Y., Poon, A., Mostovoy, Y., Hastie, A.R., Stedman, W., Anantharaman, T., Andrews, W., Zhou, X., Pang, A.W.C., Dai, H., Chu, C., Lin, C., Wu, J.J.K., Li, C.M.L., Li, J.-W., Yim, A.K.Y., Chan, S., Sibert, J., D akula, eljko, Cao, H., Yiu, S.-M., Chan, T.-F., Yip, K.Y., Xiao, M., Kwok, P.-Y., 2016. Genome-Wide Structural Variation Detection by Genome Mapping on Nanochannel Arrays. Genetics 202, 351–362.

Manrao, E.A., Derrington, I.M., Pavlenok, M., Niederweis, M., Gundlach, J.H., 2011. Nucleotide Discrimination with DNA Immobilized in the MspA Nanopore. PLoS ONE 6, e25723.

Marahrens, Y., Stillman, B., 1992. A yeast chromosomal origin of DNA replication defined by multiple functional elements. Science 255, 817–823.

Marchetti, M.A., Kumar, S., Hartsuiker, E., Maftahi, M., Carr, A.M., Freyer, G.A., Burhans, W.C., Huberman, J.A., 2002. A single unbranched S-phase DNA damage and replication fork blockage checkpoint pathway. Proc. Natl. Acad. Sci. 99, 7472–7477.

Marsh, R.C., 1978. Map location of the Escherichia coli origin of replication. Mol. Gen. Genet. MGG 166, 299–304.

Marsh, R.C., Worcel, A., 1977. A DNA fragment containing the origin of replication of the Escherichia coli chromosome. Proc. Natl. Acad. Sci. 74, 2720–2724.

Masters, M., Broda, P., 1971. Evidence for the bidirectional replications of the Escherichia coli chromosome. Nature. New Biol. 232, 137–140.

Matsui, M., Oka, A., Takanami, M., Yasuda, S., Hirota, Y., 1985. Sites of dnaA protein-binding in the replication origin of the Escherichia coli K-12 chromosome. J. Mol. Biol. 184, 529–533.

Matsuoka, S., Huang, M., Elledge, S.J., 1998. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. Science 282, 1893–1897.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., Pugh, B.F., 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res. 18, 1073–1083.

McDowall, M.D., Harris, M.A., Lock, A., Rutherford, K., Staines, D.M., Bahler, J., Kersey, P.J., Oliver, S.G., Wood, V., 2015. PomBase 2015: updates to the fission yeast database. Nucleic Acids Res. 43, D656–D661.

McGuffee, S.R., Smith, D.J., Whitehouse, I., 2013. Quantitative, Genome-Wide Analysis of Eukaryotic Replication Initiation and Termination. Mol. Cell 50, 123–135.

McNally, B., Singer, A., Yu, Z., Sun, Y., Weng, Z., Meller, A., 2010. Optical Recognition of Converted DNA Nucleotides for Single-Molecule DNA Sequencing Using Nanopore Arrays. Nano Lett. 10, 2237–2244.

Méchali, M., Kearsey, S., 1984. Lack of specific sequence requirement for DNA replication in Xenopus eggs compared with high sequence specificity in yeast. Cell 38, 55–64.

Meldal, M., Tornøe, C.W., 2008. Cu-Catalyzed Azide−Alkyne Cycloaddition. Chem. Rev. 108, 2952–3015.

Meselson, M., Stahl, F.W., 1958. The replication of DNA in Escherichia coli. Proc. Natl. Acad. Sci. 44, 671–682.

Mesner, L.D., Crawford, E.L., Hamlin, J.L., 2006. Isolating Apparently Pure Libraries of Replication Origins from Complex Genomes. Mol. Cell 21, 719–726.

Mesner, L.D., Valsakumar, V., Cieslik, M., Pickin, R., Hamlin, J.L., Bekiranov, S., 2013. Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. Genome Res. 23, 1774–1788.

Mesner, L.D., Valsakumar, V., Karnani, N., Dutta, A., Hamlin, J.L., Bekiranov, S., 2011. Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. Genome Res. 21, 377–389.

Michalet, X., Ekong, R., Fougerousse, F., Rousseaux, S., Schurra, C., Hornigold, N., van Slegtenhorst, M., Wolfe, J., Povey, S., Beckmann, J.S., Bensimon, A., 1997. Dynamic molecular combing: stretching the whole human genome for high-resolution studies. Science 277, 1518–1523.

Miotto, B., Ji, Z., Struhl, K., 2016. Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. Proc. Natl. Acad. Sci. 113, E4810–E4819.

Miyabe, I., Morishita, T., Shinagawa, H., Carr, A.M., 2009. Schizosaccharomyces pombe Cds1Chk2 regulates homologous recombination at stalled replication forks through the phosphorylation of recombination protein Rad60. J. Cell Sci. 122, 3638–3643.

Monahan, B.J., Villén, J., Marguerat, S., Bähler, J., Gygi, S.P., Winston, F., 2008. Fission yeast SWI/SNF and RSC complexes show compositional and functional differences from budding yeast. Nat. Struct. Mol. Biol. 15, 873–880.

Morishita, T., Tsutsui, Y., Iwasaki, H., Shinagawa, H., 2002. The Schizosaccharomyces pombe rad60 Gene Is Essential for Repairing Double-Strand DNA Breaks Spontaneously Occurring during Replication and Induced by DNA-Damaging Agents. Mol. Cell. Biol. 22, 3537–3548.

Mortensen, U.H., Lisby, M., Rothstein, R., 2009. Rad52. Curr. Biol. 19, R676–R677.

Mott, M.L., Berger, J.M., 2007. DNA replication initiation: mechanisms and regulation in bacteria. Nat. Rev. Microbiol. 5, 343–354.

Mukhopadhyay, R., Lajugie, J., Fourel, N., Selzer, A., Schizas, M., Bartholdy, B., Mar, J., Lin, C.M., Martin, M.M., Ryan, M., Aladjem, M.I., Bouhassira, E.E., 2014. Allele-Specific Genome-wide Profiling in Human Primary Erythroblasts Reveal Replication Program Organization. PLoS Genet. 10, e1004319.

Murray, J.M., Lindsay, H.D., Munday, C.A., Carr, A.M., 1997. Role of Schizosaccharomyces pombe RecQ homolog, recombination, and checkpoint genes in UV damage tolerance. Mol. Cell. Biol. 17, 6868–6875.

Nagata, T., 1963. The molecular synchrony and sequential replication of DNA in Escherichia coli. Proc. Natl. Acad. Sci. 49, 551–559.

Naik, A., Alzeer, J., Triemer, T., Bujalska, A., Luedtke, N.W., 2017. Chemoselective modification of vinyl DNA by triazolinediones. Angew. Chem. Int. Ed.

Neef, A.B., Luedtke, N.W., 2014. An Azide-Modified Nucleoside for Metabolic Labeling of DNA. ChemBioChem 15, 789–793.

New, J.H., Sugiyama, T., Zaitseva, E., Kowalczykowski, S.C., 1998. Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A. Nature 391, 407.

Nguyen, V.Q., Co, C., Li, J.J., 2001. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. Nature 411, 1068–1073.

Nieduszynski, C.A., 2006. Genome-wide identification of replication origins in yeast by comparative genomics. Genes Dev. 20, 1874–1879.

Nimonkar, A.V., Amitani, I., Baskin, R.J., Kowalczykowski, S.C., 2007. Single Molecule Imaging of Tid1/Rdh54, a Rad54 Homolog That Translocates on Duplex DNA and Can Disrupt Joint Molecules. J. Biol. Chem. 282, 30776–30784.

Ning, X., Guo, J., Wolfert, M.A., Boons, G.-J., 2008. Visualizing Metabolically Labeled Glycoconjugates of Living Cells by Copper-Free and Fast Huisgen Cycloadditions. Angew. Chem. Int. Ed. 47, 2253–2255.

Octobre, G., Lorenz, A., Loidl, J., Kohli, J., 2008. The Rad52 Homologs Rad22 and Rti1 of Schizosaccharomyces pombe Are Not Essential for Meiotic Interhomolog Recombination, but Are Required for Meiotic Intrachromosomal Recombination and Mating-Type-Related DNA Repair. Genetics 178, 2399–2412.

Okuno, Y., Satoh, H., Sekiguchi, M., Masukata, H., 1999. Clustered adenine/thymine stretches are essential for function of a fission yeast replication origin. Mol. Cell. Biol. 19, 6699–6709.

Orr-Weaver, T.L., 1991. Drosophila chorion genes: cracking the eggshell's secrets. BioEssays News Rev. Mol. Cell. Dev. Biol. 13, 97–105.

Osheim, Y.N., Miller, O.L., Beyer, A.L., 1988. Visualization of Drosophila melanogaster chorion genes undergoing amplification. Mol. Cell. Biol. 8, 2811–2821.

Paixao, S., Colaluca, I.N., Cubells, M., Peverali, F.A., Destro, A., Giadrossi, S., Giacca, M., Falaschi, A., Riva, S., Biamonti, G., 2004. Modular Structure of the Human Lamin B2 Replicator. Mol. Cell. Biol. 24, 2958–2967.

Pankratz, D.G., Forsburg, S.L., 2005. Meiotic S-phase damage activates recombination without checkpoint arrest. Mol. Biol. Cell 16, 1651–1660.

Pasero, P., Bensimon, A., Schwob, E., 2002. Single-molecule analysis reveals clustering and epigenetic regulation of replication origins at the yeast rDNA locus. Genes Dev. 16, 2479–2484.

Patel, P.K., Arcangioli, B., Baker, S.P., Bensimon, A., Rhind, N., 2006. DNA replication origins fire stochastically in fission yeast. Mol. Biol. Cell 17, 308–316.

Paulovich, A.G., Hartwell, L.H., 1995. A checkpoint regulates the rate of progression through S phase in S. cerevisiae in response to DNA damage. Cell 82, 841–847.

Paulovich, A.G., Margulies, R.U., Garvik, B.M., Hartwell, L.H., 1997. RAD9, RAD17, and RAD24 are required for S phase regulation in Saccharomyces cerevisiae in response to DNA damage. Genetics 145, 45–62.

Petes, T.D., Newlon, C.S., 1974. Structure of DNA in DNA replication mutants of yeast. Nature 251, 637–639.

Petryk, N., Kahli, M., d'Aubenton-Carafa, Y., Jaszczyszyn, Y., Shen, Y., Silvain, M., Thermes, C., Chen, C.-L., Hyrien, O., 2016. Replication landscape of the human genome. Nat. Commun. 7, 10208.

Petukhova, G., Stratton, S., Sung, P., 1998. Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. Nature 393, 91–94.

Petukhova, G., Sung, P., Klein, H., 2000. Promotion of Rad51-dependent D-loop formation by yeast recombination factor Rdh54/Tid1. Genes Dev. 14, 2206–2215.

Picard, F., Cadoret, J.-C., Audit, B., Arneodo, A., Alberti, A., Battail, C., Duret, L., Prioleau, M.-N., 2014. The Spatiotemporal Program of DNA Replication Is Associated with Specific Combinations of Chromatin Marks in Human Cells. PLoS Genet. 10, e1004282.

Prescott, D.M., Kuempel, P.L., 1972. Bidirectional replication of the chromosome in Escherichia coli. Proc. Natl. Acad. Sci. 69, 2842–2845.

Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J.H.J., Becker-Ziaja, B., Boettcher, J.P., Cabeza-Cabrerizo, M., Camino-Sánchez, Á., Carter, L.L., Doerrbecker, J., Enkirch, T., Dorival, I.G.-, Hetzelt, N., Hinzmann, J., Holm, T., Kafetzopoulou, L.E., Koropogui, M., Kosgey, A., Kuisma, E., Logue, C.H., Mazzarelli, A., Meisel, S., Mertens, M., Michel, J., Ngabo, D., Nitzsche, K., Pallasch, E., Patrono, L.V., Portmann, J., Repits, J.G., Rickett, N.Y., Sachse, A., Singethan, K., Vitoriano, I., Yemanaberhan, R.L., Zekeng, E.G., Racine, T., Bello, A., Sall, A.A., Faye, Ousmane, Faye, Oumar, Magassouba, N., Williams, C.V., Amburgey, V., Winona, L., Davis, E., Gerlach, J., Washington, F., Monteil, V., Jourdain, M., Bererd, M., Camara, Alimou, Somlare, H., Camara, Abdoulaye, Gerard, M., Bado, G., Baillet, B., Delaune, D., Nebie, K.Y., Diarra, A., Savane, Y., Pallawo, R.B., Gutierrez, G.J., Milhano, N., Roger, I., Williams, C.J., Yattara, F., Lewandowski, K., Taylor, J., Rachwal, P., J. Turner, D., Pollakis, G., Hiscox, J.A., Matthews, D.A., Shea, M.K.O., Johnston, A.M., Wilson, D., Hutley, E., Smit, E., Di Caro, A., Wölfel, R., Stoecker, K., Fleischmann, E., Gabriel, M., Weller, S.A., Koivogui, L., Diallo, B., Keïta, S., Rambaut, A., Formenty, P., Günther, S., Carroll, M.W., 2016. Real-time, portable genome sequencing for Ebola surveillance. Nature 530, 228–232.

Raghuraman, M.K., Winzeler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., Fangman, W.L., 2001. Replication dynamics of the yeast genome. Science 294, 115–121.

Rand, A.C., Jain, M., Eizenga, J.M., Musselman-Brown, A., Olsen, H.E., Akeson, M., Paten, B., 2017. Mapping DNA methylation with high-throughput nanopore sequencing. Nat. Methods 14, 411–413.

Rao, H., Tanpure, A.A., Sawant, A.A., Srivatsan, S.G., 2012. Enzymatic incorporation of an azide-modified UTP analog into oligoribonucleotides for post-transcriptional chemical functionalization. Nat. Protoc. 7, 1097–1112.

Rees, W.A., Yager, T.D., Korte, J., Von Hippel, P.H., 1993. Betaine can eliminate the base pair composition dependence of DNA melting. Biochemistry (Mosc.) 32, 137–144.

Reifenberger, J.G., Dorfman, K.D., Cao, H., 2015. Topological events in single molecules of E. coli DNA confined in nanochannels. The Analyst 140, 4887–4894.

Remus, D., Beall, E.L., Botchan, M.R., 2004. DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC–DNA binding. EMBO J. 23, 897–907.

Ren, X., El-Sagheer, A.H., Brown, T., 2015. Azide and trans-cyclooctene dUTPs: incorporation into DNA probes and fluorescent click-labelling. The Analyst 140, 2671–2678.

Renkawitz, J., Lademann, C.A., Jentsch, S., 2014. Mechanisms and principles of homology search during recombination. Nat. Rev. Mol. Cell Biol. 15, 369–383.

Rhind, N., 2006. DNA replication timing: random thoughts about origin firing. Nat. Cell Biol. 8, 1313–1316.

Rhind, N., Russell, P., 1998. The Schizosaccharomyces pombe S-phase checkpoint differentiates between different types of DNA damage. Genetics 149, 1729–1737.

Rhind, N., Russell, P., 2000. Checkpoints: it takes more than time to heal some wounds. Curr. Biol. 10, R908–R911.

Rhoads, A., Au, K.F., 2015. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics 13, 278–289.

Richter, D., Grün, R., Joannes-Boyau, R., Steele, T.E., Amani, F., Rué, M., Fernandes, P., Raynal, J.-P., Geraads, D., Ben-Ncer, A., Hublin, J.-J., McPherron, S.P., 2017. The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. Nature 546, 293–296.

Rieder, U., Luedtke, N.W., 2014. Alkene-Tetrazine Ligation for Imaging Cellular DNA. Angew. Chem. Int. Ed. 53, 9168–9172.

Rodriguez, J., Lee, L., Lynch, B., Tsukiyama, T., 2017. Nucleosome occupancy as a novel chromatin parameter for replication origin functions. Genome Res. 27, 269–277.

Rostovtsev, V.V., Green, L.G., Fokin, V.V., Sharpless, K.B., 2002. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. Angew. Chem. Int. Ed Engl. 41, 2596–2599.

Rowley, A., Cocker, J.H., Harwood, J., Diffley, J.F.X., 1995. Initiation complex assembly at budding yeast replication origins begins with the recognition of a bipartite sequence by limiting amounts of the initiator, ORC. EMBO J. 14, 2631.

Rupnik, A., Lowndes, N.F., Grenon, M., 2010. MRN and the race to the break. Chromosoma 119, 115–135.

Salic, A., Mitchison, T.J., 2008. A chemical method for fast and sensitive detection of DNA synthesis in vivo. Proc. Natl. Acad. Sci. 105, 2415–2420.

Sanchez, Y., Desany, B.A., Jones, W.J., Liu, Q., Wang, B., Elledge, S.J., 1996. Regulation of RAD53 by the ATM-like kinases MEC1 and TEL1 in yeast cell cycle checkpoint pathways. Science 271, 357–360.

Santa Maria, S.R., Kwon, Y., Sung, P., Klein, H.L., 2013. Characterization of the Interaction between the *Saccharomyces cerevisiae* Rad51 Recombinase and the DNA Translocase Rdh54. J. Biol. Chem. 288, 21999–22005.

Santocanale, C., Diffley, J.F., 1998. A Mec1- and Rad53-dependent checkpoint controls late-firing origins of DNA replication. Nature 395, 615–618.

Schaarschmidt, D., Baltin, J., Stehle, I.M., Lipps, H.J., Knippers, R., 2004. An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. EMBO J. 23, 191–201.

Seaton, B.L., Yucel, J., Sunnerhagen, P., Subramani, S., 1992. Isolation and characterization of the Schizosaccharomyces pombe rad3 gene, involved in the DNA damage and DNA synthesis checkpoints. Gene 119, 83–89.

Sekedat, M.D., Fenyö, D., Rogers, R.S., Tackett, A.J., Aitchison, J.D., Chait, B.T., 2010. GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. Mol. Syst. Biol. 6.

Sekimizu, K., Bramhill, D., Kornberg, A., 1987. ATP activates dnaA protein in initiating replication of plasmids bearing the origin of the E. coli chromosome. Cell 50, 259–265.

Shah, P.P., Zheng, X., Epshtein, A., Carey, J.N., Bishop, D.K., Klein, H.L., 2010. Swi2/Snf2-Related Translocases Prevent Accumulation of Toxic Rad51 Complexes during Mitotic Growth. Mol. Cell 39, 862–872.

Shim, J., Humphreys, G.I., Venkatesan, B.M., Munz, J.M., Zou, X., Sathe, C., Schulten, K., Kosari, F., Nardulli, A.M., Vasmatzis, G., Bashir, R., 2013. Detection and Quantification of Methylation in DNA using Solid-State Nanopores. Sci. Rep. 3.

Shinohara, A., Ogawa, T., 1998. Stimulation by Rad52 of yeast Rad51-mediated recombination. Nature 391, 404–407.

Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J., Timp, W., 2017. Detecting DNA cytosine methylation using nanopore sequencing. Nat. Methods 14, 407–410.

Simpson, R.T., 1990. Nucleosome positioning can affect the function of a cis-acting DMA element in vivo. Nature 343, 387–389.

Skarstad, K., Katayama, T., 2013. Regulating DNA Replication in Bacteria. Cold Spring Harb. Perspect. Biol. 5, a012922–a012922.

Smith, D.J., Whitehouse, I., 2012. Intrinsic coupling of lagging-strand synthesis to chromatin assembly. Nature 483, 434–438.

Sogo, J.M., 2002. Fork Reversal and ssDNA Accumulation at Stalled Replication Forks Owing to Checkpoint Defects. Science 297, 599–602.

Sogo, J.M., Stasiak, A., Martínez-Robles, M.L., Krimer, D.B., Hernández, P., Schvartzman, J.B., 1999. Formation of knots in partially replicated DNA molecules. J. Mol. Biol. 286, 637–643.

Song, C.-X., Clark, T.A., Lu, X.-Y., Kislyuk, A., Dai, Q., Turner, S.W., He, C., Korlach, J., 2011. Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. Nat. Methods 9, 75–77.

Song, L., Hobaugh, M.R., Shustak, C., Cheley, S., Bayley, H., Gouaux, J.E., 1996. Structure of Staphylococcal α-Hemolysin, a Heptameric Transmembrane Pore. Science 274, 1859.

Soni, G.V., Singer, A., Yu, Z., Sun, Y., McNally, B., Meller, A., 2010. Synchronous optical and electrical detection of biomolecules traversing through solid-state nanopores. Rev. Sci. Instrum. 81, 014301.

Sood, A., 2010. Technology S-Curve. Wiley Int. Encycl. Mark. 5.

Soriano del Amo, D., Wang, W., Jiang, H., Besanceney, C., Yan, A.C., Levy, M., Liu, Y., Marlow, F.L., Wu, P., 2010. Biocompatible Copper(I) Catalysts for in Vivo Imaging of Glycans. J. Am. Chem. Soc. 132, 16893–16899.

Soriano, I., Morafraile, E.C., Vázquez, E., Antequera, F., Segurado, M., 2014. Different nucleosomal architectures at early and late replicating origins in Saccharomyces cerevisiae. BMC Genomics 15, 791.

Stewart, E., Chapman, C.R., Al-Khodairy, F., Carr, A.M., Enoch, T., 1997. rqh1+, a fission yeast gene related to the Bloom's and Werner's syndrome genes, is required for reversible S phase arrest. EMBO J. 16, 2682–2692.

Stillman, B., 1993. DNA replication. Replicator renaissance. Nature 366, 506–507.

Stinchcomb, D.T., Struhl, K., Davis, R.W., 1979. Isolation and characterisation of a yeast chromosomal replicator. Nature 282, 39–43.

Stranges, P.B., Palla, M., Kalachikov, S., Nivala, J., Dorwart, M., Trans, A., Kumar, S., Porel, M., Chien, M., Tao, C., Morozova, I., Li, Z., Shi, S., Aberra, A., Arnold, C., Yang, A., Aguirre, A., Harada, E.T., Korenblum, D., Pollard, J., Bhat, A., Gremyachinskiy, D., Bibillo, A., Chen, R., Davis, R., Russo, J.J., Fuller, C.W., Roever, S., Ju, J., Church, G.M., 2016. Design and characterization of a nanopore-coupled polymerase for single-molecule DNA sequencing by synthesis on an electrode array. Proc. Natl. Acad. Sci. U. S. A. 113, E6749–E6756.

Sun, Z., Fay, D.S., Marini, F., Foiani, M., Stern, D.F., 1996. Spk1/Rad53 is regulated by Mec1-dependent protein phosphorylation in DNA replication and damage checkpoint pathways. Genes Dev. 10, 395–406.

Sung, P., 1997a. Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase. Genes Dev. 11, 1111–1121.

Sung, P., 1997b. Function of yeast Rad52 protein as a mediator between replication protein A and the Rad51 recombinase. J. Biol. Chem. 272, 28194–28197.

Sung, P., Trujillo, K.M., Van Komen, S., 2000. Recombination factors of Saccharomyces cerevisiae. Mutat. Res. Mol. Mech. Mutagen. 451, 257–275.

Suto, K., Nagata, A., Murakami, H., Okayama, H., 1999. A double-strand break repair component is essential for S phase completion in fission yeast cell cycling. Mol. Biol. Cell 10, 3331–3343.

Tabata, S., Oka, A., Sugimoto, K., Takanami, M., Yasuda, S., Hirota, Y., 1983. The 245 base-pair oriC sequence of the E. coli chromosome directs bidirectional replication at an adjacent region. Nucleic Acids Res. 11, 2617–2626.

Tanaka, K., Russell, P., 2001. Mrc1 channels the DNA replication arrest signal to checkpoint kinase Cds1. Nat. Cell Biol. 3, 966–972.

Taylor, J.H., Woods, P.S., Hughes, W.L., 1957. The organization and duplication of chromosomes as revealed by autoradiographic studies using tritium- labeled thymidine. Proc. Natl. Acad. Sci. U. S. A. 43, 122–128.

Theis, J.F., Newlon, C.S., 1997. The ARS309 chromosomal replicator of Saccharomyces cerevisiae depends on an exceptional ARS consensus sequence. Proc. Natl. Acad. Sci. 94, 10786–10791.

Thoma, F., Bergman, L.W., Simpson, R.T., 1984. Nuclease digestion of circular TRP1ARS1 chromatin reveals positioned nucleosomes separated by nuclease-sensitive regions. J. Mol. Biol. 177, 715–733.

Ticau, S., Friedman, L.J., Champasa, K., Corrêa, I.R., Gelles, J., Bell, S.P., 2017. Mechanism and timing of Mcm2–7 ring closure during DNA replication origin licensing. Nat. Struct. Mol. Biol. 24, 309–315.

Ticau, S., Friedman, L.J., Ivica, N.A., Gelles, J., Bell, S.P., 2015. Single-Molecule Studies of Origin Licensing Reveal Mechanisms Ensuring Bidirectional Helicase Loading. Cell 161, 513–525.

Timson, J., 1975. Hydroxyurea. Mutat. Res. 32, 115–132.

Tougan, T., Kasama, T., Ohtaka, A., Okuzaki, D., Saito, T.T., Russell, P., Nojima, H., 2010. The Mek1 phosphorylation cascade plays a role in meiotic recombination of *Schizosaccharomyces pombe*. Cell Cycle 9, 4688–4702.

Travers, K.J., Chin, C.-S., Rank, D.R., Eid, J.S., Turner, S.W., 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. 38, e159–e159.

Tribioli, C., Biamonti, G., Giacca, M., Colonna, M., Riva, S., Falaschi, A., 1987. Characterization of human DNA sequences synthesized at the onset of S-phase. Nucleic Acids Res. 15, 10211–10232.

Tsutsui, Y., Khasanov, F.K., Shinagawa, H., Iwasaki, H., Bashkirov, V.I., 2001. Multiple interactions among the components of the recombinational DNA repair system in Schizosaccharomyces pombe. Genetics 159, 91–105.

Urbañski, N.K., Beresewicz, A., 2000. Generation of OH initiated by interaction of Fe2+ and Cu+ with dioxygen; comparison with the Fenton chemistry. ACTA Biochim. Pol.-Engl. Ed.- 47, 951–962.

Valton, A.-L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintome, C., Riou, J.-F., Prioleau, M.-N., 2014. G4 motifs affect origin positioning and efficiency in two vertebrate replicators. EMBO J. 33, 732–746.

van den Bosch, M., Vreeken, K., Zonneveld, J.B., Brandsma, J.A., Lombaerts, M., Murray, J.M., Lohman, P.H., Pastink, A., 2001. Characterization of RAD52 homologs in the fission yeast Schizosaccharomyces pombe. Mutat. Res. Repair 461, 311–323.

van den Bosch, M., Zonneveld, J.B.M., Vreeken, K., de Vries, F.A.T., Lohman, P.H.M., Pastink, A., 2002. Differential expression and requirements for Schizosaccharomyces pombe RAD52 homologs in DNA repair and recombination. Nucleic Acids Res. 30, 1316–1324.

Varadaraj, K., Skinner, D.M., 1994. Denaturants or cosolvents improve the specificity of PCR amplification of a G +. Gene 140, 1–5.

Vas, A., Mok, W., Leatherwood, J., 2001. Control of DNA Rereplication via Cdc2 Phosphorylation Sites in the Origin Recognition Complex. Mol. Cell. Biol. 21, 5767–5777.

Vashee, S., Cvetic, C., Lu, W., Simancek, P., Kelly, T.J., Walter, J.C., 2003. Sequence-independent DNA binding and replication initiation by the human origin recognition complex. Genes Dev. 17, 1894–1908.

Vassilev, L., Johnson, E.M., 1989. Mapping initiation sites of DNA replication in vivo using polymerase chain reaction amplification of nascent strand segments. Nucleic Acids Res. 17, 7693–7705.

Vassilev, L., Johnson, E.M., 1990. An initiation zone of chromosomal DNA replication located upstream of the c-myc gene in proliferating HeLa cells. Mol. Cell. Biol. 10, 4899–4904.

Vassilev, L.T., Burhans, W.C., DePamphilis, M.L., 1990. Mapping an origin of DNA replication at a single-copy locus in exponentially proliferating mammalian cells. Mol. Cell. Biol. 10, 4685–4689.

Venditti, P., Costanzo, G., Negri, R., Camilloni, G., 1994. ABFI contributes to the chromatin organization of Saccharomyces cerevisiae ARS1 B-domain. Biochim. Biophys. Acta BBA - Gene Struct. Expr. 1219, 677–689.

Vernis, L., Piskur, J., Diffley, J., 2003. Reconstitution of an efficient thymidine salvage pathway in Saccharomyces cerevisiae. Nucleic Acids Res. 31, 120e–120.

Wallace, E.V.B., Stoddart, D., Heron, A.J., Mikhailova, E., Maglia, G., Donohoe, T.J., Bayley, H., 2010. Identification of epigenetic DNA modifications with a protein nanopore. Chem. Commun. 46, 8195.

Waltz, S.E., Trivedi, A.A., Leffak, M., 1996. DNA replication initiates non-randomly at multiple sites near the c-myc gene in HeLa cells. Nucleic Acids Res. 24, 1887–1894.

Wang, W., Hong, S., Tran, A., Jiang, H., Triano, R., Liu, Y., Chen, X., Wu, P., 2011. Sulfated Ligands for the Copper(I)-Catalyzed Azide-Alkyne Cycloaddition. Chem. - Asian J. 6, 2796–2802.

Watson, J.D., Crick, F.H., 1953a. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171, 737–738.

Watson, J.D., Crick, F.H., 1953b. Genetical implications of the structure of deoxyribonucleic acid. Nature 171, 964–967.

Wechsler, J.A., Gross, J.D., 1971. Escherichia coli mutants temperature-sensitive for DNA synthesis. Mol. Gen. Genet. MGG 113, 273–284.

Weinert, T.A., Kiser, G.L., Hartwell, L.H., 1994. Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair. Genes Dev. 8, 652–665.

Wescoe, Z.L., Schreiber, J., Akeson, M., 2014. Nanopores Discriminate among Five C5-Cytosine Variants in DNA. J. Am. Chem. Soc. 136, 16582–16587.

Williams, R.S., Williams, J.S., Tainer, J.A., 2007. Mre11–Rad50–Nbs1 is a keystone complex connecting DNA repair machinery, double-strand break signaling, and the chromatin templateThis paper is one of a selection of papers published in this Special Issue, entitled 28th International West Coast Chromatin and Chromosome Conference, and has undergone the Journal's usual peer review process. Biochem. Cell Biol. 85, 509–520.

Willis, N., Rhind, N., 2009a. Mus81, Rhp51 (Rad51), and Rqh1 form an epistatic pathway required for the S-phase DNA damage checkpoint. Mol. Biol. Cell 20, 819–833.

Willis, N., Rhind, N., 2009b. Regulation of DNA replication by the S-phase DNA damage checkpoint. Cell Div. 4, 13.

Willis, N., Rhind, N., 2010. The Fission Yeast Rad32(Mre11)-Rad50-Nbs1 Complex Acts Both Upstream and Downstream of Checkpoint Signaling in the S-Phase DNA Damage Checkpoint. Genetics 184, 887–897.

Willis, N., Rhind, N., 2011. Studying S-phase DNA damage checkpoints using the fission yeast Schizosaccharomyces pombe. Methods Mol. Biol. Clifton NJ 782, 13–21.

Willis, N.A., Zhou, C., Elia, A.E.H., Murray, J.M., Carr, A.M., Elledge, S.J., Rhind, N., 2016. Identification of S-phase DNA damage-response targets in fission yeast reveals conservation of damage-response networks. Proc. Natl. Acad. Sci. 113, E3676–E3685.

Wilmes, G.M., 2004. Interaction of the S-phase cyclin Clb5 with an "RXL" docking sequence in the initiator protein Orc6 provides an origin-localized replication control switch. Genes Dev. 18, 981–991.

Wood, V., Harris, M.A., McDowall, M.D., Rutherford, K., Vaughan, B.W., Staines, D.M., Aslett, M., Lock, A., Bahler, J., Kersey, P.J., Oliver, S.G., 2012. PomBase: a comprehensive online resource for fission yeast. Nucleic Acids Res. 40, D695–D699.

Wright, A.P.H., Maundrell, K., Shall, S., 1986. Transformation of Schizosaccharomyces pombe by non-homologous, unstable integration of plasmids in the genome. Curr. Genet. 10, 503–508.

Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M., Tackett, A., Wang, G., Hon, L.S., Fang, G., Swenberg, J.A., Xiao, A.Z., 2016. DNA methylation on N6-adenine in mammalian embryonic stem cells. Nature 532, 329–333.

Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., Aparicio, O.M., 2001. Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. Science 294, 2357–2360.

Xiao, M., Phong, A., Ha, C., Chan, T.-F., Cai, D., Leung, L., Wan, E., Kistler, A.L., DeRisi, J.L., Selvin, P.R., Kwok, P.-Y., 2007. Rapid DNA mapping by fluorescent single molecule detection. Nucleic Acids Res. 35, e16–e16.

Xu, J., Yanagisawa, Y., Tsankov, A.M., Hart, C., Aoki, K., Kommajosyula, N., Steinmann, K.E., Bochicchio, J., Russ, C., Regev, A., Rando, O.J., Nusbaum, C., Niki, H., Milos, P., Weng, Z., Rhind, N., 2012. Genome-wide identification and characterization of replication origins by deep sequencing. Genome Biol. 13, R27.

Xu, W., Aparicio, J.G., Aparicio, O.M., Tavaré, S., 2006. Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. BMC Genomics 7, 276.

Yabuki, N., Terashima, H., Kitada, K., 2002. Mapping of early firing origins on a replication profile of budding yeast. Genes Cells 7, 781–789.

Yahara, I., 1972. The origin and direction of chromosomal replication in Escherichia coli. Jpn. J. Genet. 47, 33–44.

Yamada, H., Mori, H., Momoi, H., Nakagawa, Y., Ueguchi, C., Mizuno, T., 1994. A fission yeast gene encoding a protein that preferentially associates with curved DNA. Yeast Chichester Engl. 10, 883–894.

Yang, S.C.-H., Bechhoefer, J., 2008. How Xenopus laevis embryos replicate reliably: Investigating the random-completion problem. Phys. Rev. E 78.

Yardimci, H., Loveland, A.B., van Oijen, A.M., Walter, J.C., 2012. Single-molecule analysis of DNA replication in Xenopus egg extracts. Methods 57, 179–186.

Yasuda, S., Hirota, Y., 1977. Cloning and mapping of the replication origin of Escherichia coli. Proc. Natl. Acad. Sci. 74, 5458–5462.

Yin, S., Deng, W., Hu, L., Kong, X., 2009. The impact of nucleosome positioning on the organization of replication origins in eukaryotes. Biochem. Biophys. Res. Commun. 385, 363–368.

Yoshikawa, H., Sueoka, N., 1963. Sequential replication of Bacillus subtilis chromosome, I. Comparison of marker frequencies in exponential and stationary growth phases. Proc. Natl. Acad. Sci. 49, 559–566.

Zahid, O.K., Zhao, B.S., He, C., Hall, A.R., 2016. Quantifying mammalian genomic DNA hydroxymethylcytosine content using solid-state nanopores. Sci. Rep. 6.

Zakrzewska-Czerwinska, J., Jakimowicz, D., Zawilak-Pawlik, A., Messer, W., 2007. Regulation of the initiation of chromosomal replication in bacteria. FEMS Microbiol. Rev. 31, 378–387.

Zheng, T., Jiang, H., Wu, P., 2013. Single-Stranded DNA as a Cleavable Linker for Bioorthogonal Click Chemistry-Based Proteomics. Bioconjug. Chem. 24, 859–864.

Zou, L., Stillman, B., 2000. Assembly of a complex containing Cdc45p, replication protein A, and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbf4p kinase. Mol. Cell. Biol. 20, 3086–3096.