Developing single molecule methods to study DNA replication dynamics in yeasts

Emma-Louise Heron

Sir William Dunn School of Pathology

Hertford College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy in Pathology

Word count: 48,884

Acknowledgements

I have been lucky enough to receive scientific guidance from several supervisors during my studies – Dr. Conrad Nieduszynski, Dr. Carolin Müller, Prof. Stephen Kearsey, and Prof. Chris Norbury. I am extremely grateful to them all for their support and advice. In particular, Carolin has been a huge source of inspiration, motivation, and scientific guidance since my first day in the lab. I cannot thank her enough for her unwavering patience and outstanding scientific supervision throughout the course of my DPhil, and for going well above and beyond what was required of her in every aspect of her role as my supervisor. I am also hugely grateful to Conrad for giving me a place in the group and for his amazing scientific guidance and support during my time in the lab.

I would also like to thank the other members of the group for their support and friendship, especially Dr. Dima Batrakou for teaching me so much, Ms. Phoebe Oldach for her patience and kindness when helping me with data analysis problems, and Ms. Stephanie Barker for her fantastic support as our lab manager. I have truly enjoyed working with all of the wonderful members of the lab, and also the wider department, it has been a great working environment.

I am grateful to Prof. Jordan Raff and Prof. Alison Woollard for their support as my graduate advisors, and for providing encouragement and advice when required.

Both Amanda Williams and Rebecca Busby provided unfailingly friendly and helpful support when using the Illumina NextSeq, for which I am thankful. Similarly, Michal Maj and Line Erikson always made sure that everything worked smoothly in the flow cytometry lab and were always on hand with their expertise and friendly help when necessary.

I would also like to thank Prof. Richard Gibbons and Dr. Barbara Xella for allowing me to use the Irys and Saphyr platforms, and for their help when doing so.

Finally, I would like to thank collaborators in the lab of Prof. Tony Carr for sharing the *S. pombe* strains used in this work, as well their knowledge and expertise in working with this organism. Furthermore, I am grateful to both Dr. Duncan Smith for sharing the *RNH202-AID S. cerevisiae* strain, and Dr. Takashi Kubota for sharing the *CDC9-AID* strain.

Table of Contents

ABBREVIATIONS **1**

CHAPTER 1: INTRODUCTION **4**

1.1 The mitotic cell cycle **4**

1.2 *Saccharomyces cerevisiae* as a model organism **6**

1.3 *Schizosaccharomyces pombe* as a model organism **8**

1.4 The mechanisms of genome replication **10**

1.4.1 Initiation site licensing and firing 10

1.4.2 Priming 13

1.4.3 The directionality of DNA synthesis results in semi-discontinuous replication 14

1.4.4 Elongation: roles of DNA polymerases δ and ε 15

1.4.5 Okazaki fragment maturation 17

1.4.6 Replication termination 18

1.5 Replication initiation sites **19**

1.5.1 Saccharomyces cerevisiae initiation sites 20

1.5.2 Schizosaccharomyces pombe initiation sites 21

1.6 Ensemble methods for the study of replication dynamics **23**

1.7 Single molecule and single cell methods for the study of replication dynamics **27**

1.7.1 Single cell methods in the study of replication dynamics 27

1.7.2 Single molecule methods in the study of replication dynamics 28

1.8 Objectives of this study **39**

CHAPTER 2: MATERIALS AND METHODS **40**

2.1 Reagents and enzymes 40

2.2 Growth media 40

2.2.1 Saccharomyces cerevisiae media 40

2.2.2 Schizosaccharomyces pombe media 41

2.2.3 Escherichia coli media 43

2.3 Yeast strains, plasmids and oligonucleotides 43

2.4 Yeast methods 51

2.4.1 Growth conditions 51

2.4.2 Measurement of growth rate 52

2.4.3 Measurement of cell number by haemocytometry 52

2.4.4 Producing diploids, sporulation, and tetrad dissection 53

2.4.5 Yeast transformation 54

2.4.6 Spot assay 55

2.4.7 Protein extraction 55

2.4.8 Genomic DNA extraction 56

2.4.9 Cell cycle synchronisation 64

2.4.10 Measurement of DNA content by flow cytometry 67

2.5 Biochemical techniques 68

2.5.1 SDS polyacrylamide gel electrophoresis (PAGE) 68

2.5.2 Western blot 69

2.6 Molecular genetic techniques 70

2.6.1 Polymerase chain reaction (PCR) 70

2.6.2 Cloning methods 71

2.6.3 Sanger sequencing 75

2.6.4 Agarose gel electrophoresis 75

2.6.5 Southern blotting 77

2.6.6 Densitometry analysis of gel and Southern membrane images 80
2.6.7 Fluorometric measurement of DNA concentration 81
2.6.8 Spectrophotometric measurement of DNA concentration and purity 81
2.6.9 Ethanol precipitation of DNA 81
2.6.10 Solid Phase Reversible Immobilisation (SPRI) bead purification of DNA 81
2.6.11 dUTP-Cy3 incorporation trials 82
2.6.12 Fluorescent labelling of OFs by nick translation and OF repair 83
2.6.13 Measurement of DNA molecule lengths by the Agilent 4200 TapeStation System 84
2.6.14 Measurement of Illumina NextSeq library concentration by quantitative PCR (qPCR) 85

2.7 Genomic methods 85
2.7.1 Preparation of genomic DNA for Irys or Saphyr analysis 85
2.7.2 5-Bromo-2'-deoxyuridine (BrdU) immunoprecipitation (IP) and preparation of libraries for Illumina NextSeq 550 sequencing 89
2.7.3 Illumina NextSeq 550 DNA sequencing 91
2.7.4 Preparation of ONT sequencing libraries and flow cell loading 92

2.8 Bioinformatics 94
2.8.1 BrdU-IP data analysis 94
2.8.2 D-NAscent data analysis 97
2.8.3 BNG data analysis 100

CHAPTER 3: APPLYING D-NASCENT TO THE STUDY GENOME REPLICATION IN *S. POMBE* 104

3.1 Introduction 104

3.2 Verification of G1 cell cycle synchronisation in *S. pombe* 105

3.3 Assessment of BrdU incorporation by immunoprecipitation followed by short-read sequencing 108

3.3.1 Addition of 1 μM BrdU during a G1 arrest 109
3.3.2 Addition of 0.5 μM BrdU at different time points during G1 arrest and release 114
3.3.3 Addition of 0.5 μM BrdU during a G1 arrest, followed by an HU arrest 119

3.4 Validation of molecule length by PFGE 126

3.5 Assessment of BrdU incorporation by D-NAscent 130
3.5.1 Addition of 0.5 μM BrdU during a G1 arrest, followed by an HU arrest 130
3.5.2 Addition of 1 μM BrdU during a G1 arrest, followed by an HU arrest 134
3.5.3 Addition of 1 μM BrdU during a G1 arrest 142

3.6 Summary 146

CHAPTER 4: GENERATING AND INVESTIGATING STRAINS FOR OPTICAL MAPPING OF POLYMERASE USAGE ON SINGLE MOLECULES 148

4.1 Introduction 148

4.2 Optical mapping of *S. cerevisiae* and *S. pombe* genomes using Irys 152

4.3 Using gel electrophoresis and Southern blotting to detect increased ribonucleotide incorporation 157

4.4 Generating and testing *S. pombe* strains for the polymerase-usage-based approach 161
4.4.1 Regulating RNase H2 activity using a hormone-binding domain 161
4.4.2 Transcriptional regulation of RNase H2 using the nmt81 promoter 167

4.5 Generating polymerase-usage-based strains in *S. cerevisiae* 171
4.5.1 Characterising the effect of RNH201-3xminiAID on the growth of a Pol ε mutant .. 172
4.5.2 Characterising the effect of RNH202-AID on the growth of a Pol ε mutant 177
4.5.3 Generating strains with inducible pol2-M644G 179

4.6 Summary 183

CHAPTER 5: OPTICAL MAPPING OF OKAZAKI FRAGMENTS IN SINGLE MOLECULES 184

5.1 Introduction 184

5.2 Detecting DNA fragment size decreases when DNA ligase I is depleted in *S. cerevisiae* 186

5.3 *In vitro* DNA labelling trials 192

5.4 Labelling unligated Okazaki fragments *in vitro* 195

5.5 Optical mapping of *S. cerevisiae* Okazaki fragments using Saphyr 210

5.6 Summary 222

CHAPTER 6: DISCUSSION 223

REFERENCES 231

Abstract

To ensure the faithful transmission of genetic information during the cell division cycle, the entire genome must be accurately and completely duplicated prior to segregation of the identical sister chromatids. Perturbation of this process can have severe consequences for a cell, potentially resulting in cell death or genomic instability defects that can contribute to the development of disease. Therefore, it is important to understand how such events are avoided during normal genome replication and the impact of both endogenous and exogenous stresses on the process. For decades, various experimental techniques have been developed and applied to gain insight into the dynamics of genome replication, each contributing to an ever-growing knowledge base on the subject. More recent advances have included the increasing progression towards single cell and single molecule approaches to expand the understanding of cell-to-cell heterogeneity in how replication proceeds.

The current work aimed to contribute to these efforts by developing a novel approach for the analysis of single molecule replication dynamics. It was postulated that detecting DNA polymerase usage or lagging strand synthesis by optical mapping in ultra-long single molecules would enable the evaluation of each molecule's pattern of replication. A protocol for mapping Okazaki fragments was established and tested by Bionano Genomics Saphyr analysis. Preliminary results supported the detection of labelled Okazaki fragments, showing potential for further protocol and analytical optimisation to validate this new method.

Furthermore, the experimental procedure for the application of an existing single molecule technique in a different model organism was explored. Nanopore sequencing-based detection of nucleotide analogue incorporation during replication was tested in *Schizosaccharomyces pombe*. Initial analyses indicated that the experimental system was not fully optimised for the approach and required alterations to enable the inference of replication dynamics information from long sequencing reads.

Abbreviations

2D	two dimensional
ACS	ARS consensus sequence
AID	auxin-inducible degron
APS	ammonium persulphate
ARS	autonomously replicating sequence
BNG	Bionano Genomics
bp	base pairs
BrdU	5-bromo-2'-deoxyuridine
C strand	Crick strand
CDK	cyclin-dependent kinase
ChIP	chromatin immunoprecipitation
CldU	5-chloro-2'-deoxyuridine
CMG	Cdc45-Mcm27-GINS
DDK	Dbf4-dependent kinase
DLE	Direct Labelling Enzyme
DLS	direct label and stain
D-NAscent	detecting nucleotide analogue signal on extremely long nanopore traces
DSB	double strand break
dsDNA	double stranded DNA
DTT	dithiothreitol
EDTA	ethylene dinitrilo tetraacetic acid
EMM	Edinburgh minimal medium
(h)ENT	(human) equilibrative nucleoside transporter
EtBr	ethidium bromide
FACS	fluorescence-activated cell sorting
FISH	fluorescence *in situ* hybridisation
FOV	field of view
G1/G2 phase	Gap phase 1/2
Gb	giga base pairs
GINS	Go-Ichi-Ni-San
HBD	hormone-binding domain
HMW	high molecular weight
HRP	horseradish peroxidase
HU	hydroxyurea

IAA	indole-3-acetic acid
ID	identity
IDT	Integrated DNA Technologies
IdU	5-iodo-2'-deoxyuridine
IP	immunoprecipitation
KAc	potassium acetate
kb	kilo base pairs
LB	lysogeny broth
LC	label channel
LMW	low molecular weight
M	mitosis
Mb	mega base pairs
MCM	mini-chromosome maintenance
mESC	mouse embryonic stem cell
MFA	marker frequency analysis
MOPS	3-(N-morpholino)propanesulphonic acid
MQR	Molecule Quality Report
NEB	New England Biolabs
NIB	nuclear isolation buffer
NLRS	nick, label, repair, stain
OD	optical density
OF	Okazaki fragment
ONT	Oxford Nanopore Technologies
ORC	origin recognition complex
OriDB	Origin Database
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PBST	phosphate-buffered saline + 0.1% (v/v) Tween-20
PCR	polymerase chain reaction
PFGE	pulsed-field gel electrophoresis
POI	protein of interest
pre-RC	pre-replicative complex
PU-seq	polymerase-usage sequencing
PVDF	polyvinylidene difluoride
qPCR	quantitative polymerase chain reaction
S phase	synthesis phase
SCF complex	Skp, Cullin, F-box-containing complex

SDS	sodium dodecyl sulphate
SNP	single nucleotide polymorphism
SNR	signal-to-noise ratio
SPRI	solid phase reversible immobilisation
ssDNA	single stranded DNA
(HSV)-TK	(*Herpes simplex* virus) thymidine kinase
ts	temperature-sensitive
UV	ultraviolet
W strand	Watson strand
YES	yeast extract + supplements
YPA	yeast extract-peptone-adenine
YPAD	yeast extract-peptone-adenine-dextrose
YPR	yeast extract-peptone-adenine-raffinose

Chapter 1: Introduction

1.1 The mitotic cell cycle

The mitotic cell division cycle is the essential process by which cells proliferate to produce genetically-identical daughter cells. This involves the faithful transmission of genetic information through semi-conservative replication of the genome and segregation of the resulting sister chromatids (Nurse 2000). The eukaryotic cell division cycle is divided into distinct stages (Figure 1.1), separated by critical regulatory mechanisms. These checkpoints establish whether the current cell cycle phase is complete before progression to the subsequent process and prevent such transitions when they are not appropriate and would have detrimental consequences for the cell (Murray 1994; Hartwell and Weinert 1989). During the first Gap (G1) phase, a cell begins to grow, copy its organelles, and assemble the machinery required for genome replication (Limas and Cook 2019). To progress through the cell cycle, the cell must irreversibly transition through Start (in yeast cells) (Hartwell et al. 1974), or the restriction point (in metazoa) (Zetterberg and Larsson 1985; Pardee 1974). Once this transition has taken place, the cell is committed to proceeding into the synthesis (S) phase, the stage dedicated to genome replication (Laskey, Fairman, and Blow 1989). Upon completion of S phase, the cell enters a second Gap (G2) phase with two identical copies of its genome. G2 phase separates S phase from mitosis (M) and is characterised by further cell growth and organelle duplication, ensuring sufficient cellular components for division between two new cells. A G2/M checkpoint crucially safeguards against aberrant mitotic entry, for example in the presence of DNA damage (Rhind and Russell 1998). Under the appropriate conditions, cells are able to pass through this checkpoint and begin the process of mitotic division. M phase is itself divided into several stages (Mcintosh and Koonce 1989), beginning with prophase, when the cell prepares for division by chromosome condensation, spindle assembly, and nuclear envelope breakdown (except in most yeasts) (Boettcher and Barral 2013). During metaphase, each kinetochore on sister chromatids is bound by microtubules associated with opposing spindle poles and the chromosomes are aligned centrally along the

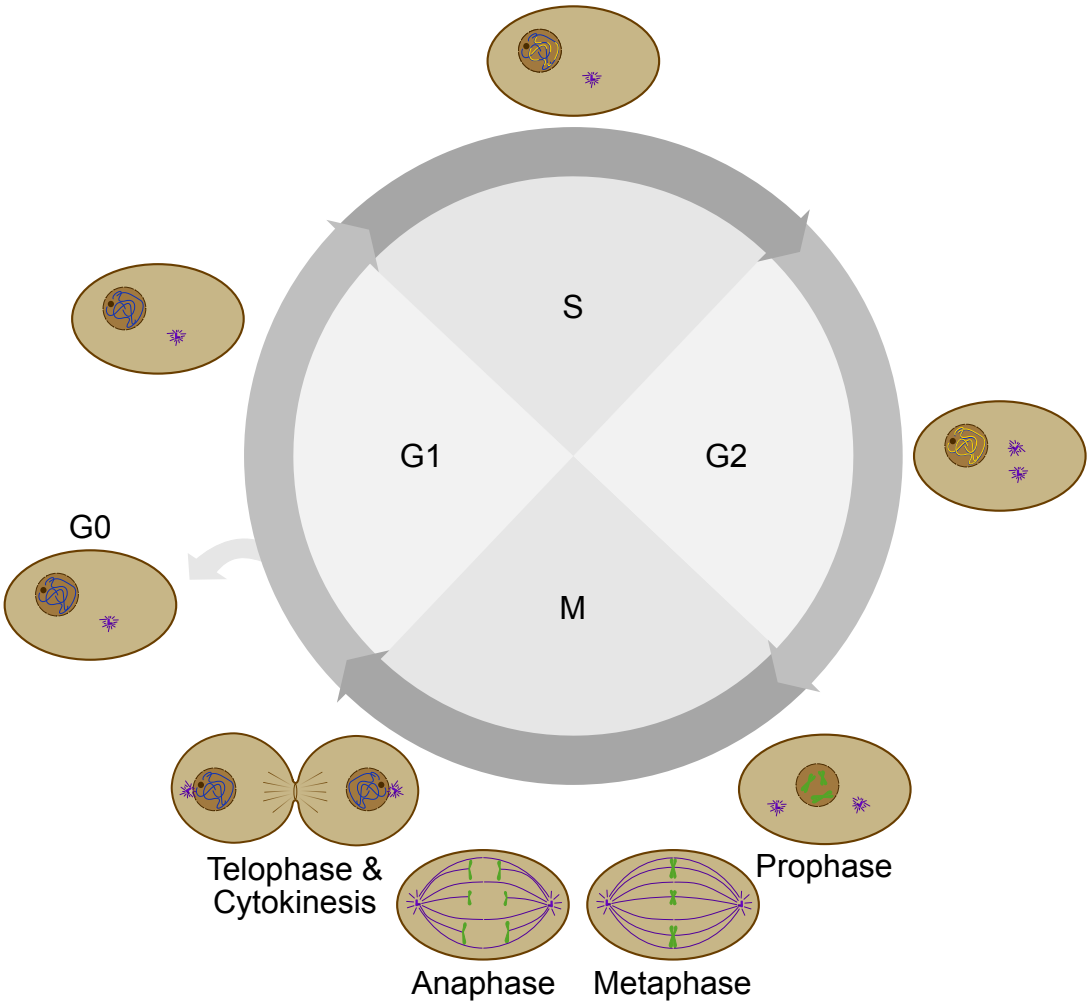


Figure 1.1*: Schematic of the mitotic cell division cycle. In G1 phase, a cell has one copy of its genome in the nucleus and one spindle body (purple). During S phase, genome replication takes place (nascent DNA is shown in yellow). By G2 phase, the cell should have two complete copies of the genome, as well as two spindle bodies. M phase commences with chromosome condensation during prophase and the breakdown of the nuclear envelope (except in most yeasts). This is followed by chromosome alignment at the metaphase plate and spindle attachment at kinetochores during metaphase. Successful attachment of all kinetochores results in the progression into anaphase, which is characterised by microtubule depolymerisation pulling sister chromatids towards opposing spindle bodies. During telophase and cytokinesis, a nuclear envelope reforms around each new genome copy, the chromosomes decondense, and cellular contents are divided into two daughter cells. The daughter cells can then re-enter G1 phase, or transition into a quiescent state, termed G0* (Gray et al. 2004).

metaphase plate (Mcintosh and Koonce 1989). It is crucial that every kinetochore is bound by a microtubule from the appropriate pole to ensure correct chromatid segregation. Therefore, a further checkpoint, the spindle-assembly checkpoint, monitors this process (Hayward, Alfonso-Pérez, and Gruneberg 2019). Progression through this checkpoint upon accurate kinetochore-microtubule interactions leads to anaphase. Microtubule disassembly during anaphase (Rath and Sharp 2011), coupled with the loss of sister chromatid cohesion (Uhlmann 2001), results in the segregation of sister chromatids towards the opposing spindle bodies. When metaphase and anaphase occur correctly, this gives rise to the separation of two complete, identical genome copies to opposite ends of the mitotic cell. Subsequently, spindle disassembly, the re-formation of a nuclear envelope around each of the two sets of segregated chromatids (where necessary), and the decompaction of the chromosomes take place during telophase (Mcintosh and Koonce 1989). Concomitant with the latter stages of mitosis, the cytoplasm of the cell is divided between two new daughter cells in the process of cytokinesis. The ultimate result is the production of two genetically-identical daughter cells from one parental cell.

1.2 *Saccharomyces cerevisiae* as a model organism

S. cerevisiae is a species of budding yeast that has been extensively studied for decades, due to the ease of culturing, its fast doubling time (in the order of 90 mins), and its genetic tractability (Sherman 2002). The term “budding” refers to its mode of mitotic cell division, during which a new ellipsoidal daughter cell buds from the mother cell. The daughter cell is initially smaller in size than the mother and must undergo further growth prior to subsequent budding from this cell (Herskowitz 1988) (Figure 1.2). Natural *S. cerevisiae* isolates are generally diploids (Nasmyth and Shore 1987), with haploids only transiently existing post meiosis, prior to mating. However, haploids can be stably maintained upon deletion or mutation of the homothallic switching endonuclease (*HO*) gene (Meiron, Nahon, and Raveh, 1995; Nasmyth, 1982). Haploid *S. cerevisiae* cells exist as one of two mating types: a or α, depending on which *MAT* allele is expressed (Duina, Miller, and Keeney 2014). The *MAT* gene product determines the expression of a- or α-specific genes in a haploid cell, as well as the expression of genes common to both a and α cells (Herskowitz 1988). Cells of each mating type secrete a specific

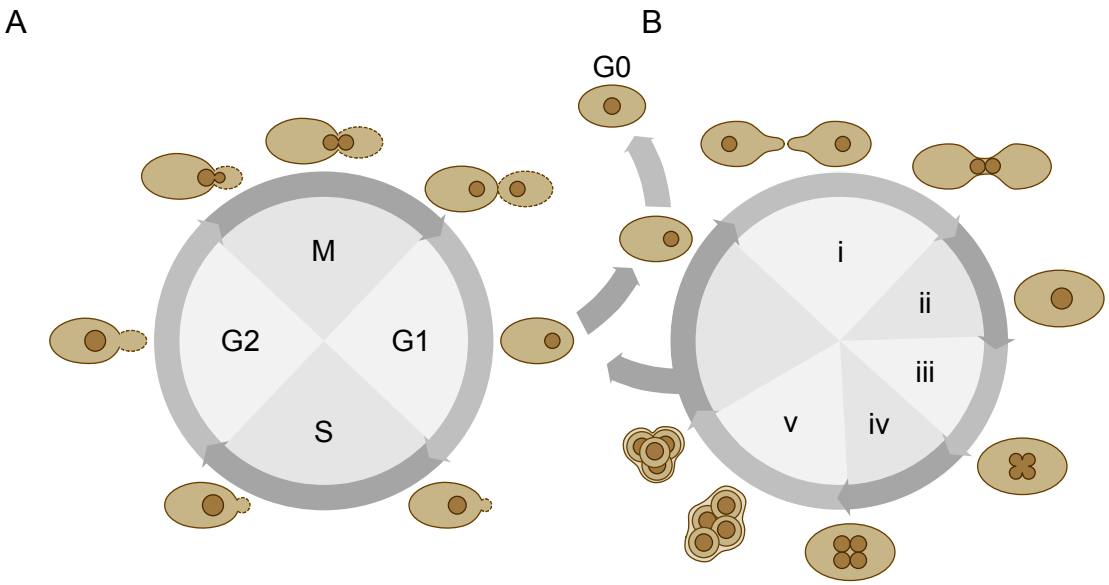


***Figure 1.2**: A schematic of the* S. cerevisiae *cell cycle. (**A**) The mitotic cell cycle. A newly-budding cell is shown by the dotted line. The budding begins during S phase and continues through G2* (Herskowitz 1988)*. The nucleus is divided in mitosis and the daughter cell buds from the mother. The daughter cell is slightly smaller than the mother. This is representative for both haploid and diploid mitosis. (**B**) Only haploid cells of opposite mating type can go on to mate by conjugation (i) to form a zygote. The zygote undergoes a round of DNA replication (ii) in preparation for meiosis I (iii) and meiosis II (iv), during which the 4C genome is divided into four haploid copies* (Neiman 2011)*. Four genetically-identical sister spores (two of each mating type) are contained within an ascus (v) before they undergo germination to yield haploid cells.*

pheromone to signal their presence to neighbouring cells. These a- or α-factor peptides are recognised by receptors on the surface of cells of the opposite mating type and induce a cell cycle arrest in G1 phase (Bardwell 2005). In addition, the receiving cell begins to grow towards the secreting cell, forming the morphology referred to as a “shmoo” (Herskowitz 1988). Mating between an a and α cell results in the formation of a zygote, with cell and nuclear fusion producing a cell in the diploid state. *MATa/MATα* diploids can be induced to sporulate by nutrient starvation, in response to which, a cell will undergo meiosis producing four haploid spores contained within an ascus (Neiman 2011). This meiotic cell cycle is extremely useful in

the laboratory context, enabling the combination and isolation of strains with required genotypes.

The *S. cerevisiae* nuclear genome comprises 16 chromosomes, and was the first eukaryotic genome to be completely sequenced (Goffeau et al. 1996). This effort identified 5,885 predicted protein-coding genes in the approximately 13.4 mega base pair (Mb) haploid genome, as well as a tandem array of ribosomal RNA genes located on chromosome XII, consisting of 100-200 repeats. The density of protein-coding genes was found to be greater than that of metazoa, with a predicted gene occurring approximately every 2 kb, and with a vast majority of these genes (96%) containing no introns. The genome sequence also provided evidence for a genome duplication event during the evolution of the species. The availability of a complete genome sequence, combined with the ease with which chromosomal sequences can be modified by homologous recombination in *S. cerevisiae* enabled the systematic deletion of nearly all of the identified genes (Giaever et al. 2002). Such work provided extraordinary insight into the function of protein-coding genes in a eukaryotic genome. The use of *S. cerevisiae* in these studies not only accelerated the development of crucial genomic approaches, but also provided predictions about the potential functions of genes in other eukaryotes, which share homology with the budding yeast genes. In addition to gene function studies, *S. cerevisiae* DNA replication initiation sites have been extensively studied and well characterised (section 1.5.1), thereby contributing to the unsurpassed annotation and understanding of a eukaryotic genome. This, in combination with the ease of experimental use and manipulation, highlights the value of *S. cerevisiae* as a model system.

1.3 *Schizosaccharomyces pombe* as a model organism

S. pombe is a yeast species that is evolutionarily distant from *S. cerevisiae*. In contrast to *S. cerevisiae*, *S. pombe* cells divide by medial fission (Forsburg and Rhind 2006) rather than budding, hence the epithet of “fission” yeast. *S. pombe* cells are rod-shaped and, during growth, extend to a critical length before division into two roughly equally-sized cells (Hoffman, Wood, and Fantes 2015). A very short G1 phase (approximately 10% of the cell cycle time

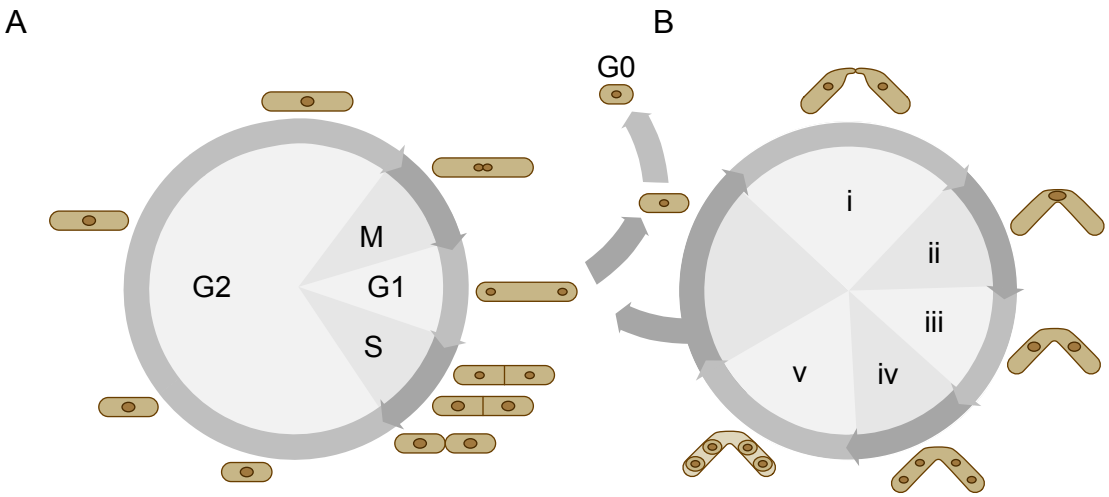


***Figure 1.3**: A schematic of the* S. pombe *cell cycle. (**A**) The mitotic cell cycle. Cell division is not completed during M or G1 phase, so binucleate cells enter S phase and undergo genome replication. Septum synthesis begins at the end of anaphase* (Gould and Simanis 1997)*, with the peak of septation coincident with mid S phase. Degradation of the primary septum is required for cell separation* (Simanis 1995)*, yielding two daughter cells concomitantly with the transition into G2 phase. The majority of the duration of the mitotic cell cycle is spent in G2 phase, when cells grow in length. After sufficient growth, cells enter mitosis, during which the nucleus divides into two, yielding binucleate cells with two identical copies of the genome. (**B**) In the meiotic cell cycle, two cells of opposite mating type (*h^+ *and* h^-*) fuse during conjugation (i). Meiotic S phase follows, resulting in a cell with a total DNA content of 4C (ii). This undergoes meiosis I to produce a binucleate 4C cell (iii). Meiosis II divides each 2C nucleus into two (iv), generating four 1C nuclei which contain the genomic information for four haploid sister spores formed during sporulation (v). Figure inspired by S. Forsburg lab website (https://dornsife.usc.edu/pombenet/fission-yeast-cell-cycle/).*

(Forsburg and Rhind 2006)) takes place prior to the completion of cell division. This means that cells are binucleate upon entry into S phase (Figure 1.3). Cytokinesis yields two G2 phase daughter cells which then grow until nuclear division takes place in mitosis (Hoffman, Wood, and Fantes 2015). G2 phase accounts for 70% of the *S. pombe* cell cycle time (Forsburg and Rhind 2006). Similar to budding yeast, haploid *S. pombe* cells exist as one of two mating types, termed h^+ and h^- (Fantes and Hoffman 2016), which can mate to produce zygotes under

conditions of nutrient starvation. However, the resulting diploids are not generally stable and usually sporulate immediately (Hoffman, Wood, and Fantes 2015), producing four ascus-contained spores. Again, this process facilitates the isolation of strains with multiple required genotypes through mating, sporulating, and subsequent tetrad dissection and analysis.

The *S. pombe* nuclear genome is organised into three chromosomes. The total size of the genome is approximately 13.8 Mb and this has been fully sequenced and annotated, with the exception of some repetitive regions in centromeres and telomeres (Wood et al. 2002). There are currently 5,065 annotated protein coding genes in the *S. pombe* genome, of which 3,539 are conserved in humans or vertebrates (Lock et al. 2019), highlighting the relevance of research on this model system to other eukaryotes. *S. pombe* centromeres range in size from 35-110 kb (Wood et al. 2002), and so are orders of magnitude longer than the *S. cerevisiae* centromeres, which are approximately 125 bp in length (Duina, Miller, and Keeney 2014), and are a closer representative of the long, complex centromeres found in metazoa. The ribosomal RNA genes are present in two tandem arrays on chromosome III (Wood et al. 2002). A comprehensive deletion collection comprising 4,836 gene deletions has provided invaluable insight into gene function and importance for viability in this organism (Kim et al. 2010). Therefore, *S. pombe* shares many of the attributes of *S. cerevisiae* as a model eukaryote, with the additional benefit of a genome organisation that more closely resembles that of other eukaryotic systems studied to date.

1.4 The mechanisms of genome replication

1.4.1 Initiation site licensing and firing

A crucial stage of the mitotic cell division cycle is the accurate and complete duplication of a cell's genome, which takes place during S phase. Eukaryotic genomes are made up of linear chromosomes, capped at each end by repetitive telomeric sequences which form secondary nucleoprotein structures (Greider 1999; Wellinger and Zakian 2012). To ensure that S phase is not prohibitively protracted, it is necessary to initiate the process of replication from multiple sites or origins of replication along each chromosome (O'Donnell, Langston, and Stillman

2013). Furthermore, the presence of multiple potential origins safeguards against under-replication by providing the ability to initiate a new event in response to replication fork stalling (Newman et al. 2013; Ibarra, Schwob, and Méndez 2008; Ge, Jackson, and Blow 2007; Woodward et al. 2006).

The preparation for genome replication begins with the interaction between the origin recognition complex (ORC) and DNA (Figure 1.4) (Diffley et al. 1994; Diffley and Cocker 1992; Bell and Stillman 1992). Subsequent ORC-binding by cell division cycle factor 6 (Cdc6, *S. cerevisiae* gene nomenclature) primes ORC for the recruitment of Cdc10-dependent transcript 1- (Cdt1)-bound mini-chromosome maintenance complex (MCM) hexamers in G1 phase (Tanaka and Diffley 2002; Donovan et al. 1997). Mcm2-7 encircle DNA in the form of head-to-head double hexamers at sites of recruitment (Remus et al. 2009), concurrently with the release of the MCM loading factors Cdc6 and Cdt1. This forms a pre-replicative complex (pre-RC), "licensing" a replication initiation site such that it is competent for activation in the subsequent S phase (Takisawa, Mimura, and Kubota 2000).

S phase entry is concomitant with rising dumbbell forming protein 4- (Dbf4)-dependent kinase (DDK) and cyclin-dependent kinase (CDK) activities (Siddiqui, On, and Diffley 2013). DDK catalyses the phosphorylation of several MCM subunits in order to activate the complex. For example, it has been shown that DDK phosphorylation of the N-terminal domain of Mcm4 is required to relieve its auto-inhibitory function (Sheu and Stillman 2010). The additional protein factors, Cdc45, synthetically lethal with DNA polymerase B 11-1 (Dpb11-1) 3 (Sld3), and Sld7 are recruited to Mcm2-7 double hexamers (Yekezare, Gómez-González, and Diffley 2013).

Increasing S phase CDK activity leads to the phosphorylation of firing factors Sld2 and Sld3. This facilitates interactions between each of the two phosphorylated Sld proteins and Dpb11 (Zegerman and Diffley 2007; Tanaka et al. 2007). Sld2 recruitment to the pre-RC results in the co-recruitment of the Go-Ichi-Ni-San (GINS) complex and the replicative DNA polymerase ε (Muramatsu et al. 2010). Together, Cdc45, the Mcm2-7 hexamer, and the GINS complex form

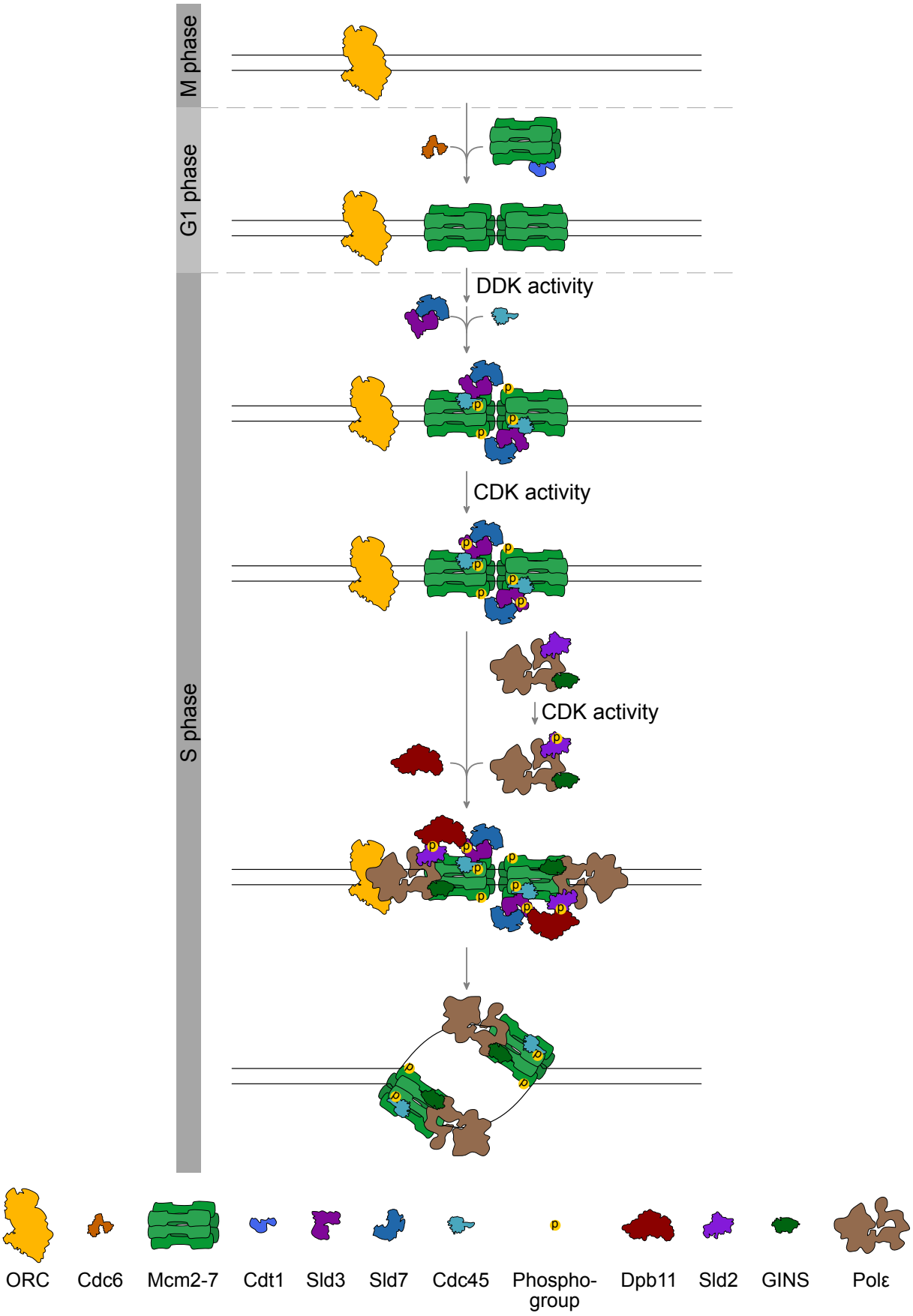


***Figure 1.4**: Schematic of pre-RC formation and origin firing. ORC can bind to potential initiation sites in the DNA during M and G1 phases* (Diffley et al. 1994). *Upon entry into G1, Cdc6 and*

Cdt1 facilitate the loading of MCM hexamers at ORC-bound sites. MCM double hexamers are loaded in the head-to-head orientation. In S phase, DDK phosphorylates MCM subunits, enabling the association of Sld3, Sld7 and Cdc45 with the MCM hexamers. Similarly, rising CDK activity results in the phosphorylation of Sld3 at the pre-RC, as well as Sld2 in complex with DNA polymerase ε and GINS. These two phospho-groups are then bound by Dpb11, bridging Sld3 and Sld2, and thereby recruiting the Sld2-DNA polymerase ε-GINS complex to the licensed initiation site. Firing takes place when the two CMG complexes bypass one another and unwind a short stretch of the parental DNA to enable template-directed priming and DNA synthesis. Figure inspired by (Yekezare, Gómez-González, and Diffley 2013).

the CMG replicative helicase (Sheu and Stillman 2006). Upon "firing", two replication forks move bidirectionally from the initiation site (Huberman and Riggs 1968), denaturing the parental strands to enable template-directed semi-conservative replication (Meselson and Stahl 1958).

Rising CDK activity triggers entry into S phase but is also inhibitory to the licensing steps. This means that replication initiation is divided between two cell cycle stages – licensing in G1, and firing in S phase. This provides a control mechanism that inhibits re-licensing of a fired origin during S phase, thereby preventing re-replication (Blow and Dutta 2005; Arias and Walter 2007). It is important for cells to avoid over-replication, as this can lead to segregation defects, ultimately resulting in aneuploidy and/or genome instability.

1.4.2 Priming

While the two principle eukaryotic replicative polymerases, δ and ε, have strand elongation activity, neither is able to initiate synthesis *de novo* (Kuchta and Stengel 2010). Thus, a primase is required to synthesise a short nucleic acid polymer, or "primer", that can then be extended by polymerase δ or ε. This role is performed by the Polα-primase complex, which comprises both an RNA primase and a DNA polymerase activity (Burgers 1998). The RNA primase produces a short RNA oligonucleotide, which is then extended by DNA polymerase α

to yield a short RNA-DNA hybrid primer (Waga and Stillman 1998). After limited polymerase α-extension, activity switches to either DNA polymerase δ or ε, which carry out the bulk of the DNA synthesis. The RNA primer at the start of each nascent strand or Okazaki fragment must be removed to preserve genome stability, as the presence of ribonucleotides in genomic DNA can result in short deletions (Nick McElhinny et al. 2010). Primer removal occurs via the strand displacement activity of DNA polymerase δ during extension of the upstream fragment (Garg et al. 2004). The displaced strand is nucleolytically degraded, thereby removing any nucleotides that are displaced and replaced by polymerase δ. It is additionally necessary to correct any errors made by polymerase α during initiator DNA synthesis because, in contrast to polymerases δ and ε, polymerase α lacks proof-reading activity (Burgers 1998) and is thus unable to remove mismatched bases it may have incorporated.

1.4.3 The directionality of DNA synthesis results in semi-discontinuous replication

DNA polymerases are only able to catalyse strand extension in the 5'-3' direction (Mitra and Kornberg 1966). Given the antiparallel structure of DNA, this means that synthesis at each replication fork is semi-discontinuous (Okazaki et al. 1968; Sakabe and Okazaki 1966), such that extended regions of single stranded DNA (ssDNA) do not persist while one strand is continuously synthesised, but rather both strands are synthesised concomitantly (Figure 1.5). The two nascent strands at a replication fork are termed the leading and lagging strands (Thömmes and Hübscher 1990). The former refers to the strand that is copied from the 3'-5' template as the DNA is unwound, which means that it can be synthesised continuously. The latter describes the strand with the 5'-3' template, which necessitates lagging strand synthesis as a series of short fragments that the replicative polymerase manufactures starting at unwound DNA slightly behind the replicative helicase. These fragments are termed Okazaki fragments (OFs) and have been shown to have a size distribution matching the nucleosome repeat pattern of 165 bp (Smith and Whitehouse 2012). Lagging strand synthesis as a series of short fragments necessitates additional processing and maturation steps to produce a complete and accurate nascent strand (section 1.4.5).

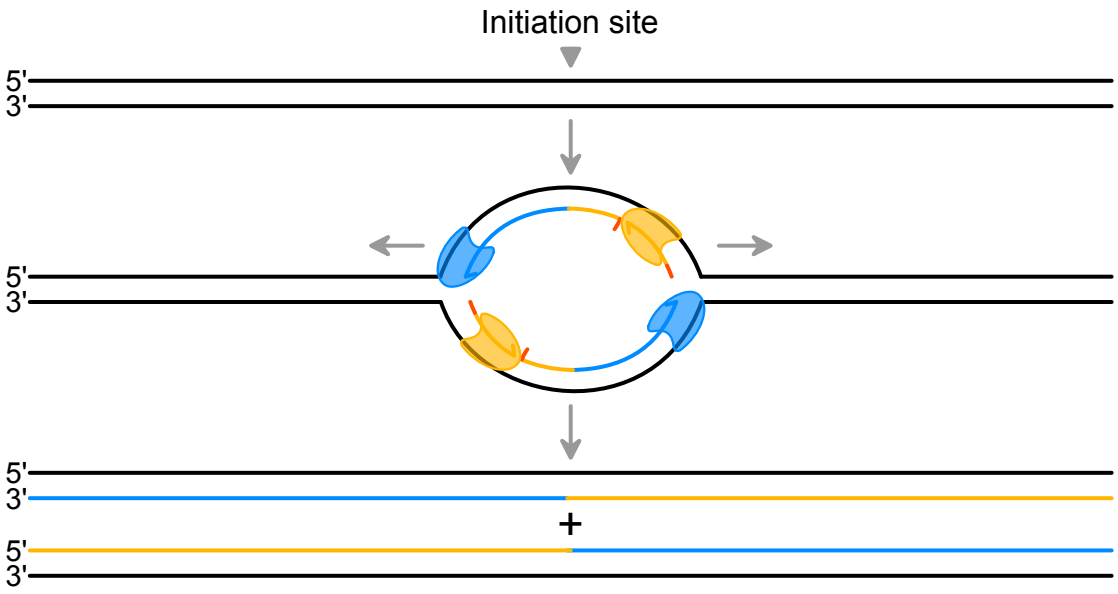


***Figure 1.5**: Schematic showing bidirectional DNA replication. Activation of an initiation site results in both a leftward- and a rightward-moving fork. At each fork, leading strand synthesis (blue) by DNA polymerase ε (section 1.4.4) is continuous due to the directionality of the polymerase activity and the orientation of the template strand. The lagging strand (yellow) is mainly synthesised by DNA polymerase δ (section 1.4.4) as a series of short OFs. Each OF is initiated by an RNA-DNA primer (red), synthesised by the Polα-primase complex (section 1.4.2). The daughter strands show the pattern of synthesis from this single initiation event.*

1.4.4 Elongation: roles of DNA polymerases δ and ε

It is widely held that during replication elongation, bulk leading strand synthesis is performed by DNA polymerase ε, while polymerase δ is responsible for lagging strand synthesis. This model was based upon the pattern of mismatches that arose during replication in *S. cerevisiae* strains expressing a mutant version of one of the replicative polymerases. For example, the substitution of methionine 644 in the polymerase catalytic domain of DNA polymerase ε was shown to result in increased levels of base mis-incorporation, despite normal exonuclease activity (Pursell et al. 2007). Measurement of the variations in T to A transversion rates in the *URA3* gene in either orientation adjacent to an active replication origin indicated that the mismatching bases were largely incorporated by polymerase ε during leading strand synthesis. In support of the conclusion that polymerase ε is the predominant polymerase involved in leading strand synthesis, similar analysis in a catalytic site mutant of polymerase δ (leucine

612 mutated to methionine) was consistent with the majority of this enzyme's activity being associated with lagging strand synthesis (Nick McElhinny, Gordenin, Stith, Burgers, and Kunkel, 2008). These results led the authors to propose a "division of labour at the eukaryotic replication fork" (Nick McElhinny et al. 2008), in which the leading strand is predominantly synthesised by DNA polymerase ε, while lagging strand synthesis is performed primarily by DNA polymerase δ (with priming by Polα-primase). More recent evidence supported this model on a genome-wide scale, by high-throughput sequencing analysis of the mutant strains (Lujan et al. 2014). Such polymerase mutants were further exploited to map strand-specific polymerase usage through the increased mis-incorporation of ribonucleotides in both *S. cerevisiae* (Clausen et al. 2015; Koh et al. 2015; Reijns et al. 2015) and *S. pombe* (Daigaku et al. 2015) (section 1.6).

However, a recent study cast uncertainty over this model for strand-specific polymerase activity (Johnson et al. 2015), leading to some controversy over the roles of the two polymerases (Johnson et al. 2016; Burgers, Gordenin, and Kunkel 2016). This study reported no strand bias for mismatches in the polymerase δ L612M mutant and instead attributed the previous observations to a role for the exonuclease domain of polymerase ε in the repair of bases misincorporated in the leading strand during synthesis by polymerase δ. In contrast, subsequent analysis of strand bias in ribonucleotide incorporation in the polymerase catalytic site mutants lent support to the division of labour model (Garbacz et al. 2018). Strand bias was lost in a strain with the polymerase δ L612M mutation and a catalytically inactive polymerase ε. Conversely, the incorporation bias was observed in a polymerase δ L612M mutant with the exonuclease activity of polymerase ε removed. These results were consistent with the division of labour model under normal conditions and suggested that polymerase ε's role at the leading strand is not simply proofreading polymerase δ's errors. However, additional work in this study implicated polymerase δ in both leading and lagging strand synthesis in regions immediately proximal to initiation sites (Garbacz et al. 2018). Two potential models were proposed to account for this observation. In the first, polymerase ε bound to CMG moves away from the initial loading site during the transition to a catalytically active form of the helicase, prior to any

primer synthesis. Therefore, polymerase ε would be unable to begin synthesis at the original protein loading site. Following exposure of the two template strands by the local helicase-induced unwinding, Polα-primase is able to synthesise the RNA-DNA primer, which is then extended by polymerase δ on both strands until it reaches the nearby CMG-polymerase ε complex, at which point polymerase ε becomes responsible for leading strand synthesis. Alternatively, but not exclusively, the observed pattern could be attributed to strand displacement synthesis by polymerase δ from the 3' end of the adjacent OF.

1.4.5 Okazaki fragment maturation

While leading strand synthesis can proceed continuously, in the absence of replication stress or fork barriers, lagging strand synthesis must be discontinuous (section 1.4.3). This necessitates RNA primer displacement and degradation for each nascent OF (Garg et al. 2004), and subsequent ligation of the nick between adjacent fragments to complete the nascent lagging strand (Johnston and Nasmyth 1978). Strand displacement by DNA polymerase δ produces short flaps which can be targeted by one of several redundant nucleases in *S. cerevisiae*: radiation sensitive factor 27 (Rad27; also known as flap endonuclease 1 (Fen1)), DNA-replication ATP-dependent helicase/nuclease 2 (Dna2), and exodeoxyribonuclease 1 (Exo1) (Liu, Hu, Wang, and Kong 2017). It was recently shown that cells expressing Rad27, in the presence or absence of Dna2 and Exo1, yielded OFs that could be ligated *in vitro* following DNA purification (Kahli et al. 2019). This suggested that Rad27 is sufficient to process the flap produced during strand displacement, generating a substrate compatible with subsequent ligation. In addition, nuclear depletion of Rad27 or deletion of *EXO1* were shown to alter the distribution OF termini positions, in a manner consistent with reduced strand-displacement activity of DNA polymerase δ. This is consistent with a model whereby strand displacement and nucleolytic degradation are mutually dependent and occur in a coordinated manner. The coupling between strand displacement and Rad27 activity was also demonstrated in *in vitro* studies of polymerase δ and Rad27 activity on artificial substrates (Stodola and Burgers 2016; Garg et al. 2004), which showed coinciding processive single nucleotide release and single nucleotide extension of a primer. Neither *EXO1* nor *RAD27* are

essential for growth in *S. cerevisiae* (Giaever et al. 2002; Reagan et al. 1995), although *rad27Δ* cells do show temperature-dependent growth defects and chromosomal instability (Reagan et al. 1995). However, deletion of *EXO1* in combination with nuclear depletion of Rad27 leads to inviability (Kahli et al. 2019). Together, these observations are consistent with some level of redundancy between the activities of Exo1 and Rad27. In contrast, *DNA2* is essential for viability (Budd and Campbell 1995) but depletion of the protein has recently been shown to have little effect on OF maturation (Kahli et al. 2019). However, it is thought to have a role in processing aberrantly long flaps under conditions where degradation by Rad27 is compromised (Ayyagari et al. 2003). Therefore, the current model for strand displacement and cleavage postulates that Rad27/Fen1 concomitantly removes nucleotides displaced by DNA polymerase δ during synthesis of a nascent OF.

It has been shown that DNA ligase cannot ligate a 3' deoxyribonucleotide to a 5' ribonucleotide *in vitro* (Ayyagari et al. 2003), therefore strand displacement must remove at least the RNA portion of the primer. Following OF primer removal, it is important that the remaining backbone nick is ligated. *CDC9* was identified as the gene encoding DNA ligase activity in *S. cerevisiae* (Johnston and Nasmyth 1978) and is essential for viability (Giaever et al. 2002). Depletion of Cdc9 protein leads to unligated OFs in *S. cerevisiae* (Smith and Whitehouse 2012), confirming this as the DNA ligase responsible for OF maturation in this species. Ligation of nicks between adjacent OFs completes the process of lagging strand maturation.

1.4.6 Replication termination

As elongating replication forks from adjacent initiation events converge, replication must be terminated to complete the nascent strands and avoid re-replication. Conversely, it is crucial that termination and replication machinery disassembly do not occur prematurely, since it is not possible to re-licence initiation sites in S phase (section 1.4.1) (Moreno and Gambus 2020). Therefore, analogous to replication initiation, termination must be tightly regulated to ensure that it takes place only when appropriate. Termination involves several distinct steps which have been investigated in a number of model systems, including reconstitution reactions,

Escherichia coli, *S. cerevisiae, Xenopus laevis* egg extract, and *Caenorhabditis elegans* embryos (Moreno and Gambus 2020).

During CMG activation at initiation sites, the helicase transitions from encircling double-stranded DNA (dsDNA) to encircling ssDNA, thereby enabling the two CMGs assembled at an initiation site to bypass each other during replication initiation (Douglas et al. 2018). This also means that, upon removal of topological stress ahead of converging forks, the two helicases are again able to bypass one another during termination (Dewar and Walter 2017). In order to disassemble the CMG from a completely replicated region, the Mcm7 subunit must be ubiquitinated by the suppressor of kinetochore protein (Skp), Cullin, F-box-containing (SCF) complex in *S. cerevisiae*. SCF bound to digs into agar 2 (Dia2), SCF^{Dia2}, forms an E3 ubiquitin ligase complex that targets Mcm7, and the deletion of *DIA2* results in CMGs persisting on the DNA into the subsequent G1 phase (Maric et al. 2014). Thus, Mcm7 ubiquitination is an essential step during CMG disassembly and replication termination. It has been proposed that the signal that triggers Mcm7 ubiquitination by SCF^{Dia2} is the transition from ssDNA to dsDNA in the CMG channel as it reaches the 5' end of the adjacent nascent strand (Dewar and Walter 2017). This would account for why such ubiquitination does not occur aberrantly during active elongation, when the helicase is translocating along ssDNA. Subsequent processing of the remaining nicks (and potential gaps) by the OF processing enzymes yields a complete nascent strand, and sister duplexes can be decatenated to conclude the process (Dewar and Walter 2017).

1.5 Replication initiation sites

Replication initiation sites have been studied in several eukaryotic model systems, from yeasts to human cells (Gilbert 2004). Intriguingly, sequence features of initiation sites do not appear to be conserved, with sites in different organisms showing distinct characteristics (Ekundayo and Bleichert 2019). The common attribute of replication initiation sites is ORC binding and pre-RC assembly, but the specificity of ORC appears to vary among different eukaryotes (MacAlpine 2016). Identification and characterisation of replication initiation sites is still

ongoing, particularly in metazoa (Urban et al. 2015; Hyrien 2015), but the current understanding of initiation sites in budding and fission yeast will be outlined below.

1.5.1 Saccharomyces cerevisiae *initiation sites*

Initiation sites have been extensively studied and well characterised in *S. cerevisiae*, and can be described as discrete origins of replication. Unlike other eukaryotic systems studied to date, *S. cerevisiae* origins contain of a specific sequence motif that is recognised and bound by ORC (Diffley and Cocker 1992; Bell and Stillman 1992). *S. cerevisiae* origins were first identified as autonomously replicating sequences (ARS), which are genome fragments that confer the ability to replicate to a vector upon insertion and transformation into yeast cells (Stinchcomb, Struhl, and Davis 1979). Efforts to pinpoint the essential sequence required for ARS activity identified a 57 bp sequence that was common to autonomously replicating plasmids (Kearsey 1983). This contained an 11 bp stretch that had a high degree of homology to a proposed ARS consensus motif. The importance of each base within such an AT-rich motif was assessed by systematic mutation and observation of the effect on ARS activity (Van Houten and Newlon 1990). In this way, the consensus sequence necessary for autonomous replication of extrachromosomal plasmids was confirmed as (A/T)TTTA(T/C)(A/G)TTT(A/T). Additional sequence elements were also found to be required for efficient ARS activity (Kearsey 1984). Such sequences are proximal to the ARS consensus sequence (ACS) and were found to unwind more readily than other sequences present in the plasmid under study (Umek and Kowalski 1988). This feature is important for replication initiation, as such helical instability facilitates DNA denaturation at origins of replication.

The potential relationship between ARS and chromosomal origins of replication was studied by two-dimensional (2D) gel electrophoresis (Huberman et al. 1988). It was found that a previously identified ARS was indeed capable of acting as an origin in yeast cells, lending weight to the hypothesis that such sequences do function as endogenous replication origins. This was subsequently investigated genome-wide, based on the prediction of origin sequences from a combination of known ARS fragments, ORC and Mcm chromatin immunoprecipitation

(ChIP) microarray data, and early-replicating copy number data (Nieduszynski, Knox, and Donaldson 2006). The majority of predicted sites were confirmed to show ARS activity. Findings from origin mapping studies in *S. cerevisiae* have been collated in an origin database (OriDB), providing information on sites that have been confirmed or predicted to function as origins (Nieduszynski et al. 2007; Siow et al. 2012).

The presence of a defined consensus sequence motif at replication initiation sites, while common to various budding yeast species (Nieduszynski, Knox, and Donaldson 2006; Theis et al. 1999), has not been identified in other eukaryotes, although certain sequence features have been associated with a higher propensity of replication initiation.

1.5.2 Schizosaccharomyces pombe *initiation sites*

An early ARS study in *S. pombe* identified nine fragments from restriction digestion of the genome that conferred ARS activity when inserted in a plasmid (Maundrell, Hutchison, and Shall 1988). These fragments ranged in length from 0.8-1.8 kb and showed increased AT-richness when compared with the genome-wide average (69-76% vs 60%). The first chromosomal initiation zone identified by 2D gel electrophoresis was a 3-4 kb region on chromosome III, but a more specific initiation site location could not be located (Zhu et al. 1992). Further analysis of fragments cloned into plasmids isolated a 2.8 kb fragment that showed initiation only within a specific region of several hundred bp, leading the authors to conclude that *S. pombe* initiation sites are localised and depend on specific sequence properties, similar to *S. cerevisiae* (Caddle and Calos 1994). This was supported by further 2D gel analysis of fragments both in autonomously-replicating plasmids and at their endogenous chromosomal locations (Wohlgemuth et al. 1994). The pattern of both active initiation and passive replication observed in the 2D gel analysis led the authors to conclude that, although *S. pombe* initiation events are localised to discrete sites on chromosomes, each of these sites may fire with a low efficiency.

In an attempt to identify sequence motifs required for ARS activity in *S. pombe*, a previously-identified highly efficient initiation site on chromosome II was subjected to a series of deletion and substitution analyses (Okuno et al. 1999). This showed that clusters of AT stretches were essential for ARS activity of this initiation site, and these were interspersed with regions that enhanced the efficiency. In support of *S. pombe* initiation sites comprising AT-rich sequences, the Orc4 subunit of the ORC in this organism was found to have nine AT-hook motifs in its N-terminal domain, which interact with AT-rich DNA sequences including a previously identified ARS (*ars1*) (Chuang and Kelly 1999; Lee et al. 2001).

Genome-wide identification of ssDNA in hydroxyurea- (HU)-arrested cells was subsequently used, both in the presence and absence of the S phase checkpoint kinase checking DNA synthesis, Cds1, as a proxy for mapping early-firing or all replication initiation sites, respectively (Feng et al. 2006). In *cds1Δ* cells, 321 peaks corresponding to enriched ssDNA were observed, 71% of which lay within 12 kb of AT-rich islands in the genome. This supported the importance of AT-richness for initiation in *S. pombe*. An alternative approach identified 401 strong initiation sites in cells that had been synchronised in G2 phase and released into the cell cycle in the presence or absence of HU, as well as a further 503 weaker signals (Heichinger et al. 2006). Furthermore, sites of pre-RC formation have been detected by Orc1 and Mcm6 ChIP (Hayashi et al. 2007). This identified 460 locations, with an average separation of 26.7 kb. Many of these pre-RC sites co-localised with regions of incorporation of the thymidine analogue 5-bromo-2'-deoxyuridine (BrdU) in HU, consistent with early initiation events. Interestingly, a deep sequencing study found that initiation site usage was more uniformly distributed following cell cycle synchronisation in G2 phase with the temperature-sensitive mutant *cdc25-22*, than following enrichment of G2 phase cells by elutriation (Xu et al. 2012). This highlighted the potential impact of the experimental conditions on the process being studied. The same work also showed a depletion of nucleosomes at the initiation sites, consistent with the inherent preclusion of nucleosome binding at poly-A sequences. Thus, it was concluded that initiation sites require two features in *S. pombe*: nucleosome depletion,

conferred by poly-A motifs, and AT-richness to facilitate ORC recruitment via the AT-hooks in Orc4.

1.6 Ensemble methods for the study of replication dynamics

It is well established that in eukaryotic cells, where the linear chromosomes are replicated from multiple initiation sites, the time in S phase at which initiation sites fire is not uniform, with some sites activating at or near the beginning of S phase, and some sites activating later (Yoshida, Poveda, and Pasero 2013; Donley and Thayer 2013). In addition, initiation sites also have the property of efficiency, which can be considered as the proportion of cells in a population that begin replication from a given initiation site in a given S phase (Raghuraman and Brewer 2010). These features of eukaryotic initiation sites contribute to species- and even cell-type-specific replication dynamics (Ryba et al. 2010). The physiological significance of variations in replication timing in different parts of the genome remain predominantly unelucidated, with several key exceptions. A recent example was the finding that the conserved early replication of the genes encoding histones H2A and H2B played an important role in ensuring maximal expression of these genes in *S. cerevisiae* (Müller and Nieduszynski 2017). Therefore, there is evidence for the biological necessity for the early replication of particular regions, which supports the observations that the replication timing patterns are not simply random.

To date, replication dynamics in model eukaryotes have principally been investigated using population-based methods, such as high-throughput sequencing, which provide an ensemble understanding of the process. There have been a number of exceptions, which are discussed in section 1.7. Several sequencing-based approaches depend on quantification of read numbers across the genome from samples enriched for S phase cells, either through fluorescence-activated cell sorting (FACS) based on overall DNA content (Müller et al. 2014; Müller and Nieduszynski 2012) or cell cycle synchronisation (Müller et al. 2014). In addition, marker frequency analysis (MFA) on exponentially growing asynchronous cells has also been used (Müller et al. 2014). The number of reads mapping to a particular genomic window indicates the relative DNA copy number of that region, ranging between 1 and 2 copies,

depending on when during S phase a sequence is replicated - values nearer to 1 correspond to regions that replicate late, and values closer to 2 indicating early replication. Thus, peaks in copy number-based replication profiles largely colocalised with previously-identified origins of replication in *S. cerevisiae*. In addition, data from cell cycle synchronisation experiments can be used to calculate the Trep; the median time in S phase at which a given locus is replicated (Müller et al. 2014). It has also been shown that the overall direction of replication forks throughout the genome can be inferred from replication profiles, and consequently, the efficiency of each origin (Hawkins et al. 2013).

Alternative approaches to the assessment of genome-wide fork direction have included specifically mapping lagging strand synthesis by sequencing OFs (McGuffee, Smith, and Whitehouse 2013; Smith and Whitehouse 2012) and determination of polymerase δ- and ε-synthesised regions through the use of increased ribonucleotide incorporation in the relevant mutants (Reijns et al. 2015; Koh et al. 2015; Daigaku et al. 2015; Clausen et al. 2015). The former of these methods involved the depletion of DNA ligase in *S. cerevisiae* through the regulation of gene expression and protein degradation using a temperature-sensitive degron (Smith and Whitehouse 2012; Smith, Yadav, and Whitehouse 2015) . This was shown to result in the S-phase-specific accumulation of DNA fragments that, upon *in vitro* ligase treatment, would decrease in abundance, consistent with the fragments corresponding to unligated OFs. Using a method that allowed the preservation of strand identity of each read enabled the genome-wide mapping of lagging strand synthesis by high-throughput sequencing (Smith and Whitehouse 2012). Importantly, complementarity was observed between the distribution of reads mapping to each strand, with sharp transitions at previously-identified origins of replication, as anticipated for bidirectional, semi-discontinuous synthesis. An origin efficiency metric was calculated from such data, providing a novel measure for the efficiency of *S. cerevisiae* origins, as well as a means of inferring the relative replication time across the whole genome (McGuffee, Smith, and Whitehouse 2013).

Four simultaneously-published studies independently developed related methods for mapping the contribution of each replicative polymerase to nascent strand synthesis in either *S. cerevisiae* (Reijns et al. 2015; Clausen et al. 2015; Koh et al. 2015) or *S. pombe* (Daigaku et al. 2015). These methods exploited known polymerase mutations that cause an increase in the aberrant incorporation of ribonucleotides into DNA during synthesis (Nick McElhinny, Kumar, et al. 2010; Daigaku et al. 2015; Lujan et al. 2013; Miyabe, Kunkel, and Carr 2011). When combined with the deletion of the RNase H2 catalytic subunit gene, *RNH201*, such ribonucleotides would persist in the genome (Rydberg and Game 2002; Nick McElhinny, Kumar, et al. 2010), as this enzyme is required for the first step in their removal from DNA-RNA hybrids. *In vitro* treatment of purified DNA that contained ribonucleotides with either an alkali, such as KOH or NaOH, or recombinant RNase HII led to backbone nicks at the phospho-ribose group. Preparation of sequencing libraries from denatured fragments of such nicked DNA by one of the approaches then allowed strand-specific mapping of the sites of ribonucleotide incorporation by each replicative polymerase mutant. This was used to determine which of the replicative polymerases was responsible for synthesising each region of each strand. Analogous to the OF-sequencing approach, sharp transitions in the signal marked strong initiation events, while more gradual transitions were indicative of more dispersed termination zones. These data provided a wealth of information, including mapping of initiation and termination sites, replication fork directionality, polymerase contribution to DNA replication, origin efficiency, and an estimate of replication timing.

An alternative approach to the study of replication dynamics is through the incorporation of nucleoside analogues (usually BrdU) during replication (Raghuraman and Brewer 2010). Yeast cells do not naturally express proteins for the thymidine salvage pathway, so they must be genetically modified to express a gene for an equilibrative nucleoside transporter (ENT) to scavenge such analogues from the medium, and a thymidine kinase (TK), to phosphorylate the imported nucleosides (Vernis, Piskur, and Diffley 2003; Sivakumar et al. 2004). Such phosphorylated analogues can be incorporated into nascent DNA by the replicative polymerases, thereby labelling sites of active DNA synthesis during the analogue pulse.

Specific antibody affinity for BrdU can be used to enrich for molecules containing the analogue and that were therefore synthesised during the BrdU treatment. This enrichment of sequences over the non-replicating regions can be detected by microarray (Viggiani, Knott and Aparicio, 2010) or more recently, by high-throughput sequencing (Haye-Bertolozzi and Aparicio 2018). A similar technique, called Repli-seq, was developed to study replication dynamics in mammalian cells (Hansen et al. 2010). In Repli-seq, a cell population is first pulsed with BrdU, then sorted by FACS to isolate fractions in different stages in S phase, as well as fractions of the population that are non-replicating (G1 and/or G2 phase). Sequencing of the purified BrdU-containing fragments can then be used to determine the point in S phase at which a given region was replicated. This has been used extensively in a variety of human and mammalian cell lines, and in the context of a wide range of biological questions, such as changes in replication timing during development, and the relationship between replication timing and genome architecture (Marchal, Sima, and Gilbert 2019).

It is evident that the methods briefly summarised here have profoundly advanced the knowledge and understanding of replication dynamics in eukaryotic cells. While mainly budding and fission yeast studies have been described, similar approaches have also been applied to metazoan cells and yielded equivalent insight (Hyrien 2015; Urban et al. 2015). One way in which to build upon such outstanding preceding work is to attempt analogous investigations into replication at the single molecule or single cell level. Ensemble methods support a model for stochastic initiation site firing which can only be directly confirmed by single cell or single molecule methods (Hawkins et al. 2013). Therefore, in order to unveil the heterogeneity between different cells and to assess potential *cis* regulation between initiation events on the same chromosome, it is necessary to develop and investigate potential single cell/molecule approaches to these questions. This would allow investigation into cell-to-cell differences in initiation site usage, and thus how the distribution of initiation events contributes to complete replication of the genome. Furthermore, such techniques could be employed to study the impact of perturbation on replication dynamics in single molecules. For example, both exogenous pressures exerted by the presence of drugs that are known to interfere with

replication, such as chemotherapeutic agents, or endogenous pressures caused by known genetic mutations, could be explored. In such cases, it is possible that a better mechanistic understanding of the mode of action of the drug or mutation might be achieved. It is important to build upon the current understanding of such questions as they are fundamental to the maintenance of genome integrity.

1.7 Single molecule and single cell methods for the study of replication dynamics

The value of single cell and single molecule methods has long been understood (Stratmann and Van Oijen 2014; Van Der Aa et al. 2013), and recent technological advances are making such techniques more readily available and applicable to a variety of biological questions. Here, a brief overview of the application of such methods to the study of genome replication will be given.

1.7.1 Single cell methods in the study of replication dynamics

Initial single cell approaches to the study of genome replication entailed cell imaging by microscopy. This has been achieved in *S. cerevisiae* through tagging of endogenous replication fork factors with fluorescent proteins to visualise replication forks in live cells or by recruitment of fluorescent proteins to repetitive arrays incorporated at a genomic locus of interest to detect when the signal increases due to duplication of the sequence (Kitamura, Blow, and Tanaka 2006). The former of these provided supporting evidence for the presence of S phase-specific replication factories in the cells, which are proposed to be clusters of replisomes at which replication takes place (Hozák and Cook 1994). Such foci were found to co-localise with sites of DNA synthesis (Kitamura, Blow, and Tanaka 2006). Subsequently, using an analogous approach, it was shown that two labelled chromosomal loci replicated in either the same or a different replication factory in different cells in a population, and that the proportion of those replicating in the same factory decreased with increasing distance between the two loci (Saner et al. 2013). The cell-to-cell variability observed in this study highlights the value of single cell methodologies for studying replication. Furthermore, stochasticity in the

replication timing of two target loci was observed using the same technique. This was also echoed in the high-throughput sequencing data (Hawkins et al. 2013), demonstrating the level of detail that can be obtained from the population-level data, as well as validating both approaches. Subsequent analysis of replication in single *S. cerevisiae* cells was used to derive an estimate for replication fork speed of approximately 2 kb/min in single cells (Dovrat et al. 2018), which was consistent with previous values given in the literature, determined using population-based methods (Sekedat et al. 2010; Raghuraman et al. 2001). Through depletion of different factors, the authors were also able to detect changes in replication timing of the target loci. For instance, a delay in replication time was observed upon depletion of the DNA ligase Cdc9, which is required for OF maturation (Johnston and Nasmyth 1978). This demonstrated the insight that could be acquired through single cell analysis of replication factor depletion or known replication mutants.

Recent developments in the analysis of replication in single cells have included sequencing-based approaches in single mammalian cells. DNA copy number variation was measured in both haploid and diploid S phase mouse embryonic stem cells (mESCs) to determine whether each locus or, in the case of the diploid, each allele had been replicated (Dileep and Gilbert 2018; Takahashi et al. 2019). Single S phase cells were sorted by FACS based on overall DNA content. Intriguingly, the replication timing appeared to be largely conserved in the individual cells analysed and matched observations from population-based methods. Thus, while single cell methods to date largely validate and confirm previous observations and models based on population approaches, the single cell methods have further potential to explore the heterogeneity of replication within a population and in cells that have been artificially perturbed. Therefore, it is likely that such methods will be crucial in expanding knowledge of replication dynamics as they are applied to more cell types and in more experimental conditions.

1.7.2 Single molecule methods in the study of replication dynamics

Several single molecule methods exist that allow the study of replication in individual molecules. Here, three main approaches will be summarised, with particular focus on Bionano

Genomics (BNG) and Oxford Nanopore Technologies (ONT), as these are directly relevant to the work undertaken in this thesis.

DNA combing

An established single molecule method for replication analysis is DNA combing (Lebofsky and Bensimon 2003). To obtain replication information from single molecules in this way, cells are treated with a nucleoside analogue, such as BrdU, 5-iodo-2'-deoxyuridine (IdU), or 5-chloro-2'-deoxyuridine (CldU). Following careful extraction, ultra-long DNA molecules are elongated on a glass surface by the flow of fluid (Bianco et al. 2012). Denaturation of the "combed" DNA reveals epitopes for immunofluorescent detection; typically a ssDNA antibody is used to target the total DNA, showing the length of the molecule, and an analogue-specific antibody is used to detect regions of incorporation. Further information can be obtained by treating cells with two analogues in succession in a double pulse experiment (Bianco et al. 2012). First, the cells are treated with, for example, IdU, for an appropriate pulse time. This is then followed by a second pulse of, for example, CldU. Distinction between the two analogues by different antibodies provides replication fork direction information, as well as the rate of fork progression. The patterns of labelling in such molecules enable the designation of replication initiation, termination, and fork stalling events. A recent advance in this approach was the maximisation of molecule lengths by an optimised combing protocol, enabling the analysis of entire *S. pombe* chromosomes, and up to 12 Mb long human DNA molecules (Kaykov et al. 2016). This allowed investigation into long-distance replication dynamics and the pattern of replication of a whole chromosome. Such analysis in *S. pombe* revealed the apparent clustering of initiation events along the chromosomes. It was also found that initiation events varied in number and location throughout S phase, with fewer active initiation sites early in S phase and an increasing rate of initiation events until the middle of S phase, when this then plateaued (Kaykov and Nurse 2015). This was a new model for how *S. pombe* chromosomal replication is organised, both spatially along each chromosome and temporally within S phase, in a clear demonstration of the value of single molecule methods.

One drawback of DNA combing approaches is the anonymity of each molecule. A region of interest in the genome can be identified using a fluorescence *in situ* hybridisation (FISH) probe complementary to the sequence of interest (Bianco et al. 2012). However, this does not provide genome-wide information and only a limited number of regions can be analysed in this way. A further shortcoming of DNA combing is the challenging methodology and the relatively low throughput. The release of novel technologies, such as BNG platforms and long-read sequencing, addresses some of these issues.

Oxford Nanopore Technologies (ONT)

ONT released the first commercial nanopore sequencer in the form of a MinION device (Loman and Watson 2015). DNA sequence information is derived from the disruption of electrical current as ssDNA migrates through protein pores embedded in an artificial membrane (Figure 1.6). The DNA perturbs the electrical signal across the pore in a manner that is dependent on the DNA sequence within the pore at the time (Jain et al. 2016). Therefore, the signal trace measured as a strand traverses the pore can be deconvoluted to determine the DNA sequence (Logsdon, Vollger, and Eichler 2020; Rang, Kloosterman, and de Ridder 2018). For most of the current ONT technology, the electrical signal is determined by the six DNA bases (6-mer) that are in the centre of the pore. The direct readout of sequence information from each strand, rather than sequencing by synthesis, means that the potential alphabet of the sequencer extends beyond the canonical four bases. For instance, it has been shown to distinguish methylated cytosine from unmethylated cytosine (Simpson et al. 2017), and crucially for replication studies, it can discriminate between BrdU and thymidine (Müller et al. 2019; Hennion et al. 2020). Additionally, nanopore technology offers the advantages of sequencing long, single molecules. These qualities of the MinION sequencer were exploited in the development of a novel method for the analysis of replication dynamics in *S. cerevisae* (Müller et al. 2019). Here, a pulse of BrdU to cells arrested in G1 phase, shortly before their release into S phase, was shown to label nascent strands around early-firing origins of replication. Furthermore, the limiting BrdU concentration was demonstrated to be sufficient to allow the inference of fork direction information through the decrease in BrdU calls as replication forks

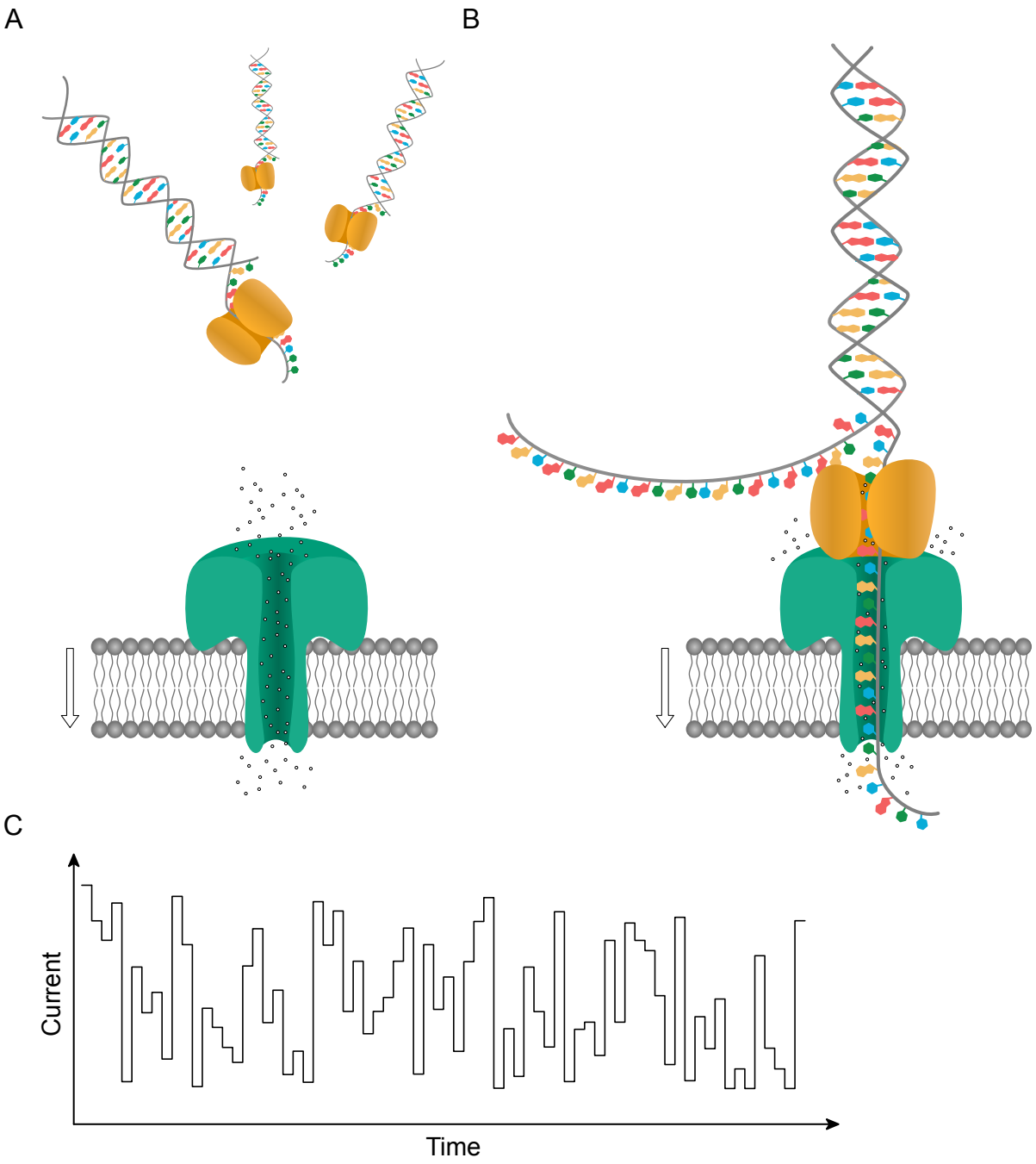


***Figure 1.6**: Schematic illustrating nanopore sequencing technology. (**A**) Protein pores (green) are embedded in an artificial membrane in MinION flow cells. The application of a voltage across the membrane drives the flow of ions through these pores (direction indicated by the arrow). DNA libraries are prepared by the ligation of an adaptor sequence to sample molecules* (Jain et al. 2016). *This adaptor sequence is bound by a motor protein (yellow) which interacts with the membrane-embedded nanopore (**B**). As the ssDNA passes through the protein pore, the ion flow is disrupted to an extent that is specific to each 6-mer that travels through the pore. (**C**) The current disruption is measured as a signal trace of current over time. ONT has released software that can de-convolute the signal traces into the DNA sequence.*

migrated away from initiation events, thereby exhausting the supply of the analogue to each fork. This work therefore described a novel protocol for detecting nucleotide analogue signal currents on extremely long nanopore traces (D-NAscent), which can theoretically be applied to investigate genome replication in any cell type that is able to (or can be genetically modified to) salvage BrdU from the medium. Thus, it represented a new method that can be recruited in the study of genome replication in long, individual molecules of known sequence.

Bionano Genomics (BNG) technology

Optical mapping approaches have provided important, complementary information to DNA sequencing, by mapping particular target sequences throughout the genome along ultra-long molecules. Originally, such methods involved treatment of combed DNA molecules with a restriction endonuclease to produce double-strand breaks (DSBs) at specific recognition sequences (Neely, Deen, and Hofkens 2011). An intercalating DNA stain enabled the visualisation of the individual DNA molecules, in which the DSBs would appear as gaps in the signal. Fluorescence microscopy was used to collect images of multiple DNA molecules, and these could then be used to construct a consensus map by aligning the patterns of restriction enzyme-derived DSBs. In contrast to many high-throughput sequencing technologies, which produce short-read data, optical mapping datasets consist of ultra-long individual DNA molecules (hundreds of kb up to Mb). Contigs generated from optical mapping data can provide scaffolds for the alignment of sequencing data, thereby offering context to the reads and contributing to more complete genome maps.

Recent advances in optical mapping methodologies include the use of nicking endonucleases or modified methyltransferases to provide the genomic location information pattern (Figure 1.7) (Neely, Deen, and Hofkens 2011). This means that ultra-long DNA molecules remain intact, rather than being cleaved, thereby enabling greater automation of DNA elongation and visualisation by electro-kinetically driving molecules through nanochannels. These developments have resulted in greater versatility in the applications of optical mapping, including the analysis of DNA replication.

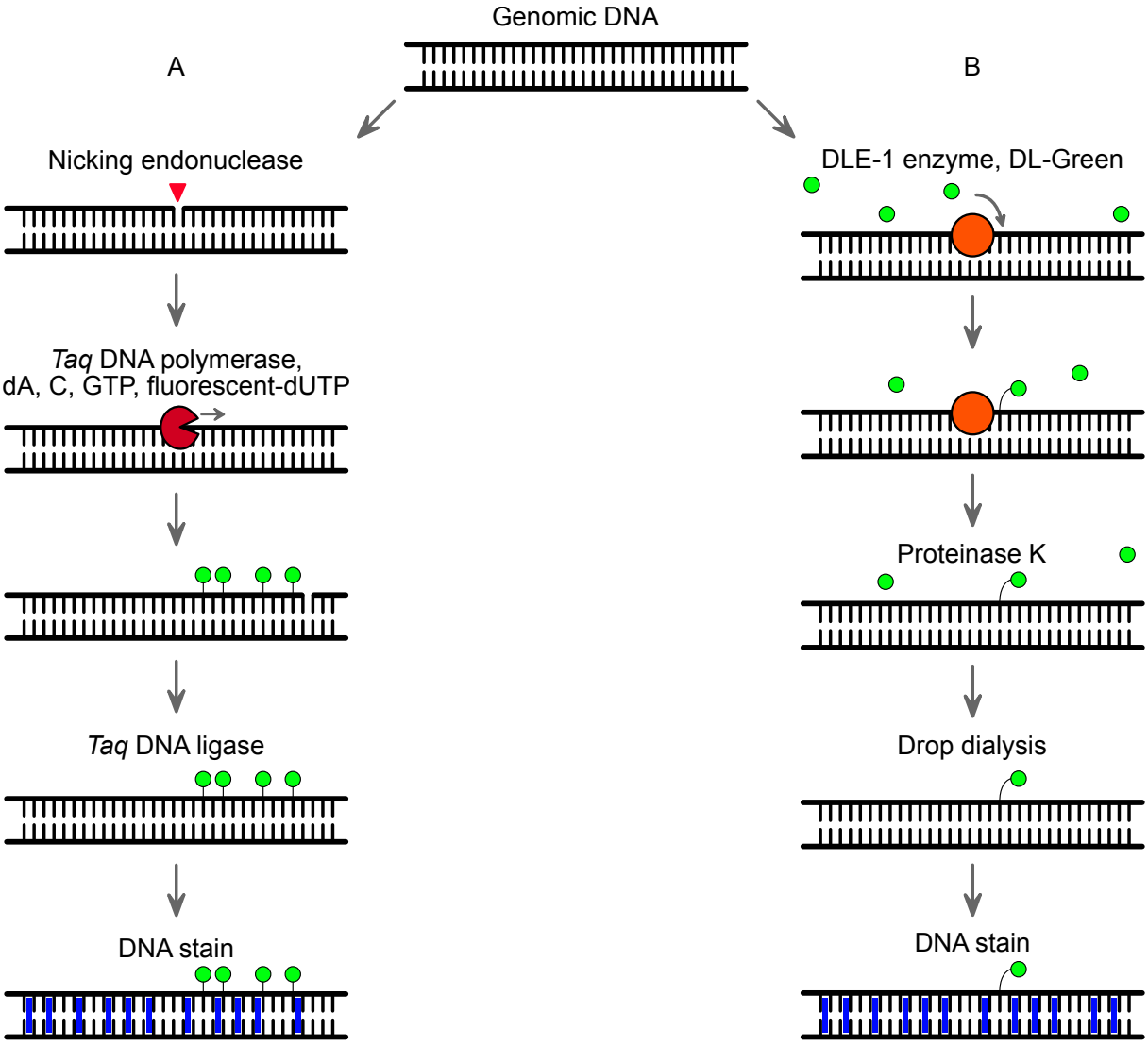


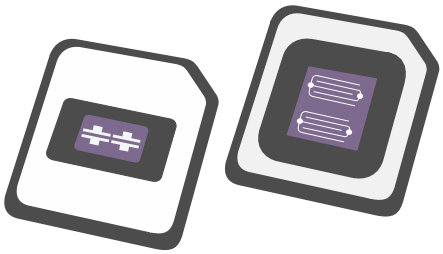
***Figure 1.7**: Schematic showing strategies for labelling genomic DNA for optical mapping. (**A**) Nick-translation to incorporate a fluorescently-labelled nucleotide at specific recognition sequence sites. This is achieved by first treating the DNA with a nicking endonuclease that has a recognition sequence that occurs at the appropriate frequency in the target genome. The enzyme cleaves the backbone of one strand of the double helix, thus producing a substrate for nick-translation. For the nick-translation reaction, the DNA is treated with a DNA polymerase capable of strand displacement and endonuclease, and/or 5'-to-3' exonuclease activity to remove DNA ahead of the nick. During the reaction, dATP, dGTP, dCTP, and fluorescently-labelled dUTP are incorporated by the DNA polymerase. Due to the low processivity of the polymerase under these conditions, the nick is translated only approximately 150 bp. Nicks are then repaired by a DNA ligase so that the DNA backbone is once again intact. DNA is stained with an intercalating dye to enable the visualisation of each*

*molecule. (**B**) An alternative approach involves treatment of the DNA with a modified methyltransferase (DLE-1) and fluorescent cofactor (DL-Green). Similar to restriction enzymes, the transferase recognises short specific sequences, at which the fluorophore is attached to the DNA. The enzyme can then be digested by proteinase K, and unattached fluorophore removed by drop dialysis. As in (A), the DNA is stained for visualisation.*

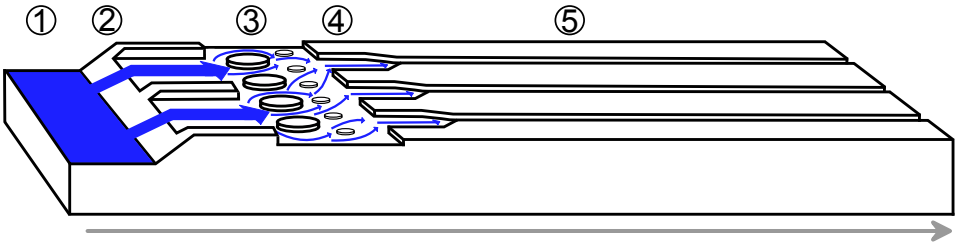
BNG released a commercial system incorporating such improvements in 2012 (Lam et al. 2012). This Irys technology combined DNA isolation via agarose plugs with a nick-labelling approach (Figure 1.7 A), followed by automated DNA elongation and visualisation to produce high genome coverage by ultra-long molecules. This has since been superseded by the analogous Saphyr and direct labelling enzyme (DLE) system. The components of the DLE reaction are proprietary, but it is likely to involve a modified methyltransferase that attaches a fluorescent moiety at its recognition sequence (Figure 1.7 B). This is advantageous over the nick-labelling method because it does not require the induction of nicks in the DNA backbone, and so preserves strand integrity and circumvents so-called “fragile sites” (where two proximal nick sites on opposite strands result in a DSB), thereby improving molecule lengths. Furthermore, false positives caused by label incorporation at nick sites independent of the nicking enzyme are avoided.

In both Irys and Saphyr systems, labelled and stained DNA is loaded into a flow cell on a chip, and the chip is then placed in the Irys or Saphyr machine. An electrical current is applied such that molecules are driven electro-kinetically through the flow cell (Figure 1.8). Flow cells are designed to detangle and separate molecules, enabling the elongation and visualisation of individual molecules in nanochannels. During a run, many thousands of molecules can be visualised across the nanochannel array, producing giga bases (Gb) of data. As the images are captured, a digital representation of each molecule is produced based on its staining and labelling, providing so-called “Molecules” information that details the length of each molecule and its label positions. Currently, these systems are capable of detecting three wavelengths – one is used for the intercalating stain, the second is for the genomic location labelling, leaving

Load DNA onto chip

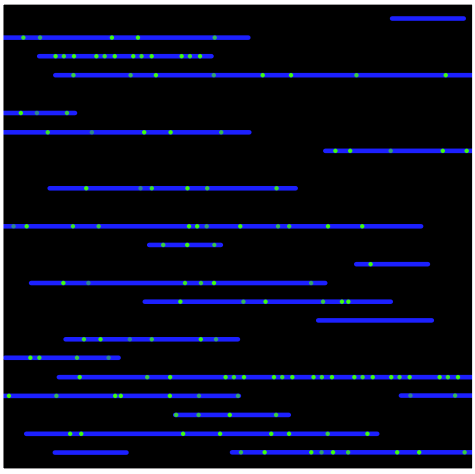


Run DNA through nanochannels



Current

Image



Molecules

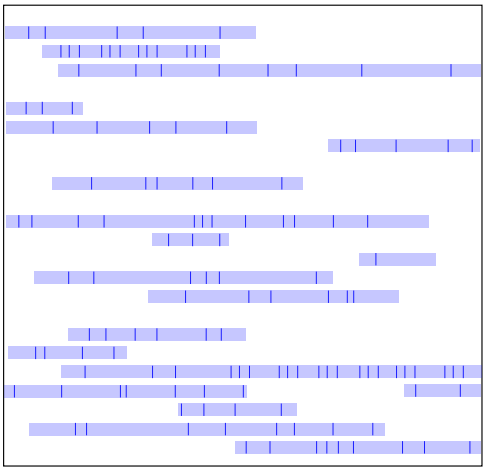


***Figure 1.8**: Schematic illustrating the BNG Irys and Saphyr systems. DNA is first loaded into inlet and outlet sections of a flow cell. Irys chips contain two flow cells, while Saphyr chips can contain two or three. This enables multiple samples to be run simultaneously in the same chip. Once the flow cell has been placed into the respective machine and the run has started, the DNA (blue) undergoes the following steps: 1 – molecules are concentrated at a lip in the flow cell, 2 – a subset of the molecules are then "spiked" over the lip by a brief increase in voltage, while the rest of the molecules remain at the lip ready for the next spike, 3 – the molecules are "bumped" towards the pillar region which begins to detangle the molecules and directs them into 4 – the microchannels, here the molecules begin to adopt a more elongated structure and are funnelled into the nanochannels in 5. Individual molecules travel through the nanochannels and images are captured by a fluorescent microscope. The raw images are automatically converted into "Molecules" information by the system, which includes details about each molecule's length and label positions.*

a third label available for other analyses. These frequently involve labelling a second recognition site to improve mapping capabilities (Yuan, Chung, and Chan 2020). This has enabled the mapping of large genomes, with long repetitive elements, such as the human genome (Wong et al. 2020).

The Molecules information can be analysed using BNG software that allows users to generate *de novo* genome assemblies from their data (Figure 1.9 A), and align the molecules or the assembly to another sample or a reference, if available. This is the intended use of the technology and enables the detection of large chromosomal rearrangements, insertions, deletions, and inversions (Yuan, Chung, and Chan 2020). Alternatively, individual molecules can be aligned to a reference to obtain mapping information about each molecule for further downstream analysis (Figure 1.9 B).

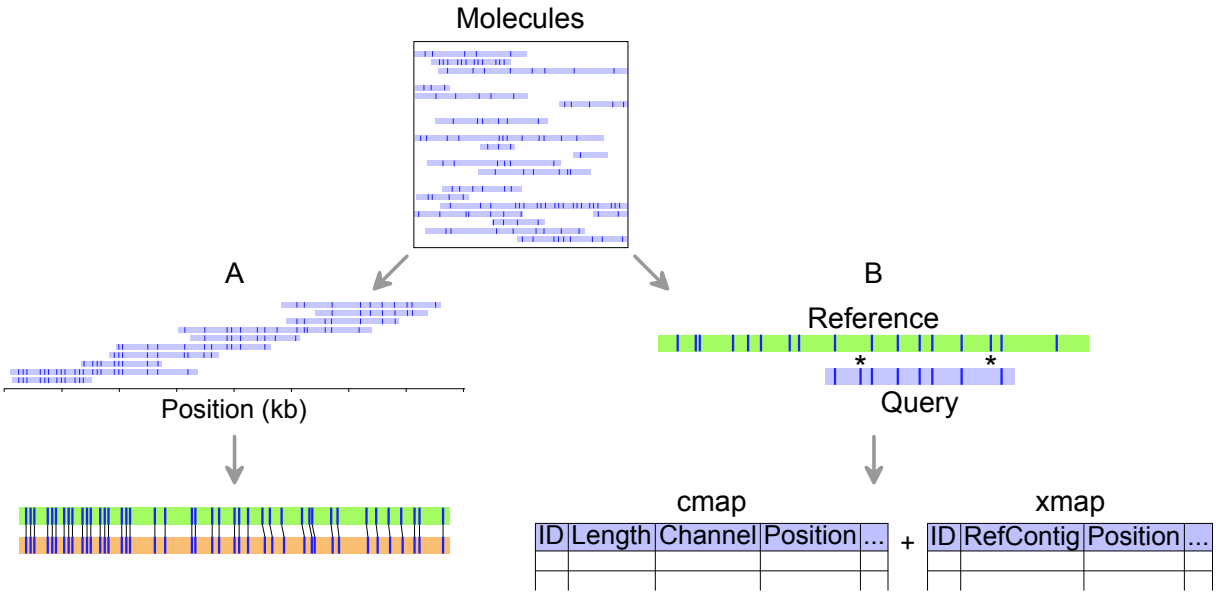


***Figure 1.9**: (**A**) BNG data analysis usually involves generating a* de novo *genome assembly by aligning molecules (blue) to produce contigs (orange). Each contig can be aligned to a reference map (green), if available. Such maps can be made by* in silico *digestion of a reference sequence with the appropriate nicking endonuclease or DLE-1. The resulting reference map provides the anticipated location of each label, based on the available sequence information. (**B**) Alternatively, it is possible to align single molecules to the reference map. This is useful if further downstream analysis, that does not require genome assembly, is necessary. The result is outputted in several file types, including an xmap, which details molecule-to-reference alignment information, and a cmap, which contains label position information for each molecule.*

The availability of a second label means that this technology can be used to study DNA modifications such as methylation (Sharim et al. 2019) or other epigenetic marks (Gabrieli et al. 2018a), DNA damage (Liu et al. 2019), or genome duplication, as discussed below.

In a study published in 2016, single DNA molecules from U2OS osteosarcoma cells were shown to migrate hydrodynamically through nanochannels in custom-made chips (Lacroix et al. 2016). Tracts of nascent DNA were visualised either by incorporation of Cy3-labelled dUTP during replication, or by an increase in fluorescence signal intensity at replication bubbles upon staining with SYTOX-Orange. While this pioneering research demonstrated the feasibility of

applying nanochannel technology to the study of DNA replication, several difficulties remained unresolved. Firstly, when using dual-colour microscopy to detect YOYO-1 stained molecules followed by visualisation of nascent DNA by dUTP-Cy3 labelling, the Cy3 signal was much reduced. This meant that it was very challenging to record the information from both the YOYO-1 staining and the Cy3 labelling, and so the Cy3 tracts could not be placed in the context of the whole molecule. Secondly, no mapping information was collected from the molecules, and therefore the location of each replication tract in the genome remained unknown. Thirdly, the use of custom-made nanochannel arrays produced in-house, while innovative, meant that the approach was unlikely to be adopted by other laboratories.

More recent work has involved the use of BNG's Irys or Saphyr technologies to study genome replication (Klein et al. 2017; De Carli et al. 2018; Wang et al. 2020). These approaches also entailed fluorescent nucleotide incorporation by replicative polymerases during replication, either of λ DNA in *Xenopus* egg extracts (De Carli et al. 2018) or in HeLa cells (Wang et al. 2020) followed by genomic location labelling and BNG analysis. In particular, the latter study illustrated the extremely high coverage that can be acquired from the Irys and Saphyr platforms, even for the human genome. It also demonstrated the value of such long single molecule data, since insight into inefficient initiation events in human cells was obtained. A limitation of this approach is that it requires transfection of fluorescently-labelled nucleotides into cells, and so is not applicable to species or cell-types where this is not possible. Furthermore, it only allows a snapshot of a small window of S phase through cell cycle synchronisation and nucleotide pulse protocols, so only the earliest replicating regions could be analysed in this way. The authors demonstrated the application of the approach to an asynchronous culture to infer replication dynamics and even fork direction data based on the intensity of fluorescent signal. However, as with the synchronised cells, only a very small fraction of the molecules had replication label signal. While this did allow the authors to calculate the efficiency of each initiation event, it also decreased the proportion of the total data that provided direct replication information.

1.8 Objectives of this study

This work aimed to build upon the methods described above to investigate a protocol for the application of D-NAscent to *S. pombe* and to develop a novel method for genome replication dynamics analysis by optical mapping. The former project was based on the novel D-NAscent approach developed in *S. cerevisiae* (Müller et al. 2019). Here, the goal was to transfer this method to *S. pombe* for the analysis of replication in this organism. This required the optimisation of a cell cycle synchronisation protocol, in which BrdU could be added to the culture during a G1 arrest to allow incorporation into nascent DNA following release into S phase. Furthermore, the appropriate BrdU concentration and DNA extraction protocol had to be investigated.

The second project that will be described aimed to exploit the ability of nick-translation reactions to incorporate a fluorescently-labelled nucleotide into DNA. The principle behind this project was to isolate DNA molecules with nicks marking replication that could then be labelled in this way. Two approaches were explored towards this aim: first, the construction of yeast strains that would have increased ribonucleotide incorporation by one of the replicative polymerases for a single S phase to allow *in vitro* nicking by RNase HII for subsequent labelling. Second, the characterisation of a protocol involving an *S. cerevisiae* strain in which DNA ligase levels could be depleted through use of a degron to yield molecules with unligated Okazaki fragments.

Ultimately, the purpose of this work was to expand the available single molecule methods for the study of replication to improve understanding of the stochasticity of this process within a population.

Chapter 2: Materials and Methods

2.1 Reagents and enzymes

Routinely-used chemicals were obtained from Merck/Sigma-Aldrich, Thermo Fisher Scientific, Qiagen, Roche, Formedium, BD Biosciences, and Amersham International. All restriction and nicking endonucleases, the *Taq* DNA polymerase used for nick-translation, Phusion High-Fidelity DNA polymerase, and *Taq* DNA ligase were acquired from New England Biolabs (NEB). For standard PCRs, BIOTAQ DNA polymerase from Bioline was used. Oligonucleotides were obtained from Integrated DNA Technologies (IDT). Zymolyse-20T and -100T were purchased from the Seikagaku Corporation. Aminoallyl-dUTP-Cy3 and aminoallyl-dUTP-ATTO-647N were procured from Jena Bioscience. Reagents specific to the Bionano Genomics (BNG), Illumina, or Oxford Nanopore Technologies (ONT) protocols were obtained from the respective companies.

2.2 Growth media

All media were autoclaved at 117 °C for 15 min.

2.2.1 Saccharomyces cerevisiae *media*

Yeast extract-Peptone-Adenine-Dextrose (YPAD) medium

YPAD broth and agar plates were made either using pre-mixed powder purchased from Formedium (CCM1005 or CCM0510, respectively), or using the following recipe: 1% yeast extract, 2% Bacto peptone, 2% dextrose, 0.005% adenine, pH adjusted to 6.0-6.5 with 1 M HCl. To produce solid medium for plates, 2% agar was added to this broth.

For positive selection of strains with the *KAN* resistance gene, G-418 antibiotic was added to a final concentration of 400 µg/mL, having allowed the medium to cool to 65 °C or less after autoclaving.

For assessment of cell growth on YPAD in the presence of the auxin indole-3-acetic acid (IAA), a 0.5 M solution of IAA in absolute ethanol was diluted to the indicated concentration using

absolute ethanol and spread on plates using glass beads. The surface was allowed to dry prior to the addition of cells.

Yeast extract-Peptone-Adenine (YPA) medium

For growth of strains in an alternative sugar source, YPA medium was made as described for YPAD, with the omission of dextrose from the recipe. A final concentration of 2% of the indicated sugar was added from a filter-sterilised 20% solution to the YPA, once cooled to 65 °C or less after autoclaving.

-ura synthetic dropout medium

For selection of strains with the *URA3* genotype, growth on medium lacking uracil was assessed. Plates lacking uracil were made as follows: pre-mixed synthetic defined medium with 2% glucose from Formedium (CSM0105) was supplemented with -URA powder (DCS0281) to a final concentration of 0.8 g/L.

Potassium acetate (KAc) medium

To induce sporulation, diploids were incubated on KAc plates, with the following composition: 2% KAc, 0.22% yeast extract, 0.05% dextrose, pH adjusted to 7.0 with 1 M HCl, 0.087% COM dropout powder from Formedium (DCS0031), 2.5% agar.

2.2.2 Schizosaccharomyces pombe *media*

Yeast extract + supplements (YES)

For growth on rich medium: 0.5% yeast extract, 3% glucose, 0.0225% each of histidine, leucine, adenine, uracil, and lysine, pH adjusted to 5.6 with 1 M HCl. To produce solid medium for plates, 2% agar was added to this broth.

For positive selection of strains with the *NAT* resistance gene, ClonNAT (Werner BioAgents) was added to a final concentration of 200 µg/mL, once the medium had cooled to 65 °C or less following autoclaving.

To allow normal RNase H2B localisation in *rnh202-HBD* transformants, YES plates were supplemented with 125 nM β-oestradiol from a 10 mM stock (dissolved in absolute ethanol) by dilution in H_2O and spreading on plates with glass beads. The surface was allowed to dry prior to the addition of cells.

Malt extract medium

To induce sporulation, diploids were incubated on malt extract plates, with the following composition: 3% malt extract, 0.0225% each of histidine, leucine, adenine, and uracil, pH adjusted to 5.5, 2% agar.

Edinburgh minimal medium (EMM) + supplements

Used for culturing strains that had to be grown in the absence of thiamine (strains with *loxP-ura4+-nmt81-lox2272-rnh201* genotype). Broth was made using 3.23% EMM broth pre-mixed powder from Formedium (PMD0202), 0.0225% each of leucine, adenine, and uracil. Solid medium made using 4.92% EMM agar pre-mixed powder from Formedium (PMD0101), 0.0225% each of leucine and adenine ± 0.0225% uracil. In experiments where expression of *loxP-ura4+-nmt81-lox2272-rnh201* was to be repressed, thiamine was added to EMM broth to a final concentration of 15 µM.

EMM + glutamate

Used for mating and sporulating strains that had to be grown in the absence of thiamine (strains with *loxP-ura4+-nmt81-lox2272-rnh201* genotype): 4.8% EMM glutamate agar pre-mixed powder from Formedium (PMD2101), 0.0225% each of leucine, adenine, and uracil.

EMM, low dextrose

Used to culture cells for transformation: 1.73% EMM broth low dextrose powder pre-mixed from Formedium (PMD0601), 0.0225% each of leucine and uracil.

2.2.3 Escherichia coli *media*

Lysogeny broth (LB) medium

LB plates and broth were purchased pre-made by departmental staff using Luria Agar Granulated [Miller's LB agar] (Melford, L24022) or Luria Broth Granulated [Miller's LB broth] (Melford, L24400), respectively.

LB plates and broth were supplemented with antibiotics for positive selection and liquid culturing of transformants expressing the relevant resistance gene as follows:

- Ampicillin: 100 µg/mL
- Chloramphenicol: 25 µg/mL
- Kanamycin: 30 µg/mL

For blue/white screening of pGEM-T Easy-derived plasmid transformants, LB_{amp} plates were supplemented with 50 µL of ChromoMax IPTG/X-gal from Fisher BioReagents (BP4200-10), which were spread on the plates using glass beads. The surface was allowed to dry prior to the addition of cells.

2.3 Yeast strains, plasmids and oligonucleotides

Table 2.1*:* S. cerevisiae *strains used in this study. All* S. cerevisiae *strains were derivatives of W303.*

Strain	Genotype	Description	Source
AUY116	*MATα, RAD5, BUD4, leu2, ura3, trp1, ade2, his3*	Control and parental WT haploid strain	The Replication Group
EHY69	*pol2-M644G, MATα*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation AUY116 transformed with pCA278 and repair construct (PCR amplified from yTAK037 DNA with SB004+AC0038) Isolate 1	This study
EHY70	*pol2-M644G, MATα*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation AUY116 transformed with pCA278 and repair construct (PCR amplified from yTAK037 DNA with SB004+AC0038)	This study

		Isolate 2	
EHY71	*URA3-pADH1-OsTIR1, rnh201::rnh201-3xminiAID, pol2-M644G, RAD5, BUD4, leu2, trp1, ade2, his3, MATa*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation RNase H2 catalytic subunit tagged with 3xminiAID Offspring of SBY014 and EHY69 cross isolated by tetrad analysis	This study
EHY86	*URA3-pGAL1-10-Os.TIR1, RAD5, BUD4, leu2, trp1, ade2, his3, MATa*	Tir1 under inducible *GAL1-10* promoter T7107 transformed with *Stu*I-linearised BYP7434	This study
EHY88	*rnh201-3xminiAID, URA3-pGAL1-10-Os.TIR1, RAD5, BUD4, leu2, trp1, ade2, his3, MATa*	Tir1 under inducible *GAL1-10* promoter RNase H2 catalytic subunit tagged with 3xminiAID EHY86 was transformed with pSB001, p414-TEF1p-Cas9-CYC1t, and repair construct amplified from BYP7430 with SB007 and SB008	This study
EHY94	*pol2-M644G, RAD5, BUD4, ura3, leu2, trp1, ade2, his3, MATa*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation Offspring of EHY69 and EHY88 cross isolated by tetrad analysis	This study
EHY96	*rnh201-3xminiAID, RAD5, BUD4, ura3, leu2, trp1, ade2, his3, MATa*	RNase H2 catalytic subunit tagged with 3xminiAID Offspring of EHY69 and EHY88 cross isolated by tetrad analysis	This study
EHY98	*pol2-M644G, rnh201-3xminiAID, RAD5, BUD4, ura3, leu2, trp1, ade2, his3, MATa*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation RNase H2 catalytic subunit tagged with 3xminiAID Offspring of EHY69 and EHY88 cross isolated by tetrad analysis	This study
EHY100	*URA3-pGAL-Os.TIR1, pol2-M644G, rnh201-3xminiAID, RAD5, BUD4, URA3, leu2, trp1, ade2, his3, MATa*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation RNase H2 catalytic subunit tagged with 3xminiAID Tir1 under inducible *GAL1-10* promoter Offspring of EHY69 and EHY88 cross isolated by tetrad analysis	This study
EHY106	*POL2, ura3-1::pGAL-pol2-M644G-3xFLAG-6xHIS-tTDH1-URA3, MATa*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation under inducible *GAL* promoter at *URA3* locus, with *POL2* at endogenous locus	This study

		T7107 transformed with *Not*I-linearised pEDH21 to integrate *pGAL-pol2-M644G-3xFLAG-6xHis-tTDH1* at the *URA3* locus	
EHY110	*POL2, ura3-1::pGAL-pol2-M644G-3xFLAG-6xHIS-tTDH1-URA3, rnh201::kanMX, MATa*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation under inducible *GAL* promoter at *URA3* locus, with *POL2* at endogenous locus RNase H2 catalytic subunit knockout by integration of a KanMX resistance cassette at this locus POY006 transformed with *Not*I-linearised pEDH21 to integrate *pGAL-pol2-M644G-3xFLAG-6xHis-tTDH1* at the *URA3* locus	This study
EHY124	*POL2, ura3-1::pGAL-pol2-M644G-3xFLAG-6xHIS-tTDH1-URA3, MATa*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation under inducible *GAL* promoter at *URA3* locus, with *POL2* at endogenous locus EHY106 was crossed with AUY116, then offspring of the sought genotype were crossed with T7107 Isolate 1 of second cross	This study
EHY125	*POL2, ura3-1::pGAL-pol2-M644G-3xFLAG-6xHIS-tTDH1-URA3, MATa*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation under inducible *GAL* promoter at *URA3* locus, with *POL2* at endogenous locus EHY106 was crossed with AUY116, then offspring of the sought genotype were crossed with T7107 Isolate 2 of second cross	This study
EHY127	*POL2, ura3-1::pGAL-pol2-M644G-3xFLAG-6xHIS-tTDH1-URA3, rnh201::kanMX, MATa*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation under inducible *GAL* promoter at *URA3* locus, with *POL2* at endogenous locus RNase H2 catalytic subunit knockout by integration of a KanMX resistance cassette at this locus EHY110 was crossed with AUY116, then offspring of the sought genotype were crossed with T7107 Isolate 1 of second cross	This study

EHY128	*POL2, ura3-1::pGAL-pol2-M644G-3xFLAG-6xHIS-tTDH1-URA3, rnh201::kanMX, MATa*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation under inducible *GAL* promoter at *URA3* locus, with *POL2* at endogenous locus RNase H2 catalytic subunit knockout by integration of a KanMX resistance cassette at this locus EHY110 was crossed with AUY116, then offspring of the sought genotype were crossed with T7107 Isolate 2 of second cross	This study
EHY130	*ura3-1::pADH1-Os.TIR1-9MYC-URA, RNH202-AID-9MYC-HIS, pol2-M644G, MATa*	Tir1 under constitutive *ADH1* promoter RNase H2B subunit tagged with full-length AID DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation Offspring of EHY69 and yDS1138 cross isolated by tetrad analysis Isolate 1	This study
EHY131	*ura3-1::pADH1-Os.TIR1-9MYC-URA, RNH202-AID-9MYC-HIS, pol2-M644G, MATa*	Tir1 under constitutive *ADH1* promoter RNase H2B subunit tagged with full-length AID DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation Offspring of EHY69 and yDS1138 cross isolated by tetrad analysis Isolate 2	This study
POY006	*rnh201::KanMX, MATa*	RNase H2 catalytic subunit knockout by integration of a KanMX resistance cassette at this locus T7107 modified by homologous recombination	Phoebe Oldach of The Replication Group
POY015	*pol2-M644G, rnh201::KanMX, MATa*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation RNase H2 catalytic subunit knockout by integration of a KanMX resistance cassette at this locus Offspring of POY006 and EHY69 cross isolated by tetrad analysis	Phoebe Oldach of The Replication Group
SBY014	*URA3-pADH1-OsTIR1, rnh201::rnh201-3xminiAID, RAD5, BUD4, leu2, trp1, ade2, his3, MATa*	Tir1 under constitutive *ADH1* promoter RNase H2 catalytic subunit tagged with 3xminiAID	Stephanie Barker of The Replication Group

T7107	*MATa, RAD5, BUD4, leu2, ura3, trp1, ade2, his3*	Control and parental WT haploid strain	Tanaka lab
TKY375	*cdc9-IAA17::kanMX, URA3::pRS306-pGAL1-10-OsTIR1 codon plus, [pRS313 (HIS3)], bar1Δ::hisG ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 RAD5, MATa*	DNA ligase 1 tagged with full-length AID Tir1 under inducible *GAL1-10* promoter Bar1 protease deleted Derivative of YK402	Kubota lab
yDS1138	*ura3-1::ADH1-OsTIR1-9MYC-URA, RNH202-AID-9MYC-HIS, RAD5, can1-100, his3-11,15, leu2-3,112, trp1-1, MATa*	Tir1 under constitutive *ADH1* promoter RNase H2B subunit tagged with full-length AID	Smith lab
YK402	*bar1Δ::hisG ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 RAD5, MATa*	Bar1 protease deleted Wild-type control for TKY375	Kubota lab

***Table 2.2**:* S. pombe *strains used in this study.*

Strain	Genotype	Description	Source
309	*ura4-D18, leu1-32, h$^-$*	Control and parental WT haploid strain	Carr lab
AW1138	*loxP-ura4+-nmt81-lox2272-rnh201, ura4-D18, leu1-32, h$^+$*	RNase H2 catalytic subunit under *nmt81* promoter	Carr lab
EHY57	*rnh202-HBD-natMX, ura4-D18, leu1-32, h$^-$*	RNase H2B subunit tagged with hormone binding domain 309 transformed with linearised pEDH5 Isolate 1	This study
EHY58	*rnh202-HBD-natMX, ura4-D18, leu1-32, h$^-$*	RNase H2B subunit tagged with hormone binding domain 309 transformed with linearised pEDH5 Isolate 2	This study
EHY59	*rnh202-HBD-natMX, cdc6-L591G, ura4-D18, leu1-32, h$^+$*	DNA polymerase δ catalytic subunit mutant for increased ribonucleotide incorporation RNase H2B subunit tagged with hormone binding domain Isolate from cross and tetrad dissection of EHY57 and IM995	This study
EHY60	*rnh202-HBD-natMX, cdc20-M630F, ura4-D18, leu1-32, h$^-$*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation RNase H2B subunit tagged with hormone binding domain Isolate from cross and tetrad dissection of EHY57 and IM997	This study
EHY67	*cdc6-L591G, loxP-ura4+-nmt81-lox2272-rnh201, ade6-704, ura4-D18, leu1-32*	DNA polymerase δ catalytic subunit mutant for increased ribonucleotide incorporation RNase H2 catalytic subunit under *nmt81* promoter Isolate from cross and tetrad dissection of AW1138 and IM996	This study

EHY68	*cdc20-M630F, loxP-ura4+-nmt81-lox2272-rnh201, ade6-704, ura4-D18, leu1-32*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation RNase H2 catalytic subunit under *nmt81* promoter Isolate from cross and tetrad dissection of AW1138 and IM998	This study
IM643	*rnh201::KanMX, ade6-704, ura4-D18, leu1-32, h$^-$*	RNase H2 catalytic subunit knockout by integration of a KanMX resistance cassette at this locus	Carr lab
IM995	*cdc6-L591G, ade6-704, ura4-D18, leu1-32, h$^+$*	DNA polymerase δ catalytic subunit mutant for increased ribonucleotide incorporation	Carr lab
IM996	*cdc6-L591G, ade6-704, ura4-D18, leu1-32, h$^-$*	DNA polymerase δ catalytic subunit mutant for increased ribonucleotide incorporation	Carr lab
IM997	*cdc20-M630F, ade6-704, ura4-D18, leu1-32, h$^+$*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation	Carr lab
IM998	*cdc20-M630F, ade6-704, ura4-D18, leu1-32, h$^-$*	DNA polymerase ε catalytic subunit mutant for increased ribonucleotide incorporation	Carr lab
YDP66	*cdc10-m17, leu1::[leu1-adh:hENT1], his7$^+$, adh1::hsv-TK, smt-0, leu1-32, ura4-D18, ade6-704, h$^-$*	BrdU uptake and phosphorylation cassette and temperature sensitive *cdc10* allele for G1 arrest at restrictive temperature (36 °C)	Carr lab

***Table 2.3**: Plasmids used in this study.*

Plasmid	Description	Source
BYP7430	3xminiAID for C-terminus tagging Mini-degron is a codon-optimised *IAA17* from *Arabidopsis thaliana* With KanMX resistance cassette for positive selection in *S. cerevisiae*	The Replication Group
BYP7434	*pGAL1-10-Os.TIR1* for integration at *URA3* in *S. cerevisiae*	The Replication Group
p414-TEF1p-Cas9-CYC1t	Human optimised *Streptococcus pyogenes* Cas9 Expression in yeast driven from *TEF1* promoter With *TRP1* for positive selection in *S. cerevisiae*	The Replication Group
p637	pFA6a-HBD-natMX6 β-oestradiol hormone binding domain (HBD)	Kearsey lab
pCA278	pCRCT with gRNA against *S. cerevisiae POL2* to create M644G mutation With *URA3* for positive selection in *S. cerevisiae*	Carolin Müller of The Replication Group
pDB18	gRNA structural component plasmid into which gRNAs could be cloned following *Bsa*I digestion With *URA3* for positive selection in *S. cerevisiae* Used to produce pEDH7	Dzmitry Batrakou of The Replication Group
pEDH5	pFA6a-*rnh202*$^+$-HBD-natMX6 *S. pombe rnh202*$^+$ tagged with the β-oestradiol HBD	This study

	EH246 and EH247 were used to amplify *rnh202*$^{+}$ from *S. pombe* genomic DNA The amplicon was digested with *Sal*I and *Pac*I and inserted into *Sal*I- and *Pac*I-cut p637	
pEDH7	pDB18 modified to encode gRNA targeting 3' end of *DBF4* in *S. cerevisiae* EH262 and EH263 were annealed and ligated into *Bsa*I-cut pDB18 With *URA3* for positive selection in *S. cerevisiae*	This study
pEDH9	*pol2-M644G* fragment 2 (chrXIV: 148,473 – 150,637) in pGEM-T Easy	This study
pEDH10	*pol2-M644G* fragment 4 (chrXIV: 151,017 – 152,083) in pGEM-T Easy	This study
pEDH11	*pol2-M644G* fragment 5 (chrXIV: 152,079 – 155,881) in pGEM-T Easy	This study
pEDH19	*pol-M644G* part plasmid assembly	This study
pEDH21	Cassette assembly of *pGAL1-pol2-M644G-3xFLAG-6xHis-tTDH1* in pYTK096	This study
pGEM-T Easy	PCR product cloning system	Promega
pSB001	Encodes gRNA targeting 3' region of *RNH201* in *S. cerevisiae*	Stephanie Barker of The Replication Group
pYTK001	Part Plasmid Entry Vector for Yeast Toolkit (Lee et al. 2015), confers chloramphenicol resistance in *E. coli*	Addgene
pYTK030	Type 2 Part Vector encoding *pGAL1* for Yeast Toolkit (Lee et al. 2015), confers chloramphenicol resistance in *E. coli*	Addgene
pYTK060	Type 4a Part Vector encoding *3xFLAG-6xHis* for Yeast Toolkit (Lee et al. 2015), confers chloramphenicol resistance in *E. coli*	Addgene
pYTK066	Type 4b Part Vector encoding *tTDH1* for Yeast Toolkit (Lee et al. 2015), confers chloramphenicol resistance in *E. coli*	Addgene
pYTK096	Pre-assembled *URA3* integration cassette vector for Yeast Toolkit (Lee et al. 2015), confers kanamycin resistance in *E. coli*	Addgene

***Table 2.4**: Oligonucleotides used in this study.*

Name	Sequence (5' to 3')	Description
AC0019	ACAAAGGCAGACATAGTACGCTAAT	Targets *S. cerevisiae RNH201* Used with SB011 to distinguish between the untagged and 3xminiAID-tagged genes
AC0036	TTTGTAGTCATGATGTTTGGGTACA	*S. cerevisiae POL2 (M644)* allele-specific primer for genotyping
AC0037	TTTGTAGTCATGATGTTTGGGTATC	*S. cerevisiae pol2-M644G* allele-specific for genotyping
AC0038	CGATGCAGTCGCTACGTATTA	*S. cerevisiae POL2*-targeting primer for repair construct production and genotype confirmation
AC0064	TGTCGACTTGAAGCTCCCAC	Used to sequence verify pEDH19
AC0071	GCTGCAAATGGCTCGGAAAA	Used to sequence verify pEDH9

AC0073	TGTACAGAAGACTATCACAGGAAACT	Used to sequence verify pEDH11
AC0074	TCCTCTATTTCAATACATGCTAATCCA	Used to sequence verify pEDH11
Alu29	CGTCAAGACTGTCAAGGAGG	Used to sequence verify pEDH5
CA007	ACAAAAACCTGCAGGAAACG	Targets *S. cerevisiae URA3* Used with CA037 and CA938 to genotype *URA3-pGAL-Os.TIR1* strains
CA033	GCACATCAATGCATCAATCC	Targets *S. cerevisiae URA3* Used with CA037 to genotype *URA3-pGAL-Os.TIR1* strains
CA035	GGATGAGTAGCAGCACGTTC	Targets *S. cerevisiae URA3* Used with MH012 to genotype *URA3-pGAL-Os.TIR1* strains
CA037	ACGCATGAAATCCTTCATTTG	Targets *S. cerevisiae URA3* Used with CA007 and CA033 to genotype *URA3-pGAL-Os.TIR1* strains
CA788	ACGCACGGAGGAGAGTCTT	Targets *pGAL1-10* Used with DB081 to genotype *URA3-pGAL-Os.TIR1* strains
CA938	GCAGGTGCCAGAGATGATG	Targets *Os.TIR1* gene Used with CA007 to genotype *URA3-pGAL-Os.TIR1* strains
CN2169	CAGCTGAAGCTTCGTACGC	Used to sequence verify pEDH5
DB081	GACCTACACCGAACTGAGATAC	Targets *ColE1* bacterial origin of replication Used with CA788 to genotype *URA3-pGAL-Os.TIR1* strains
DB117	AGTGCCCTCTTGGGCTAGCGG	Used to sequence verify pEDH7
EH216	TTTAAATCAAATCTTCCATGCG	Used to make probe template and single-stranded probe for denaturing Southern blot
EH217	GATGCCAGACCGTAATGACAAAA	Used to make probe template and single-stranded probe for denaturing Southern blot
EH246	TTTGTCGACCAGGGAAAGATATTCATACTACCC	Used with EH247 to amplify *rnh202*+ from *S. pombe* DNA
EH247	CCCTTAATTAACTTTTTTGTAAAGAATGAAGATATTTTCGTC	Used with EH246 to amplify *rnh202*+ from *S. pombe* DNA
EH248	AGGCTGTTCTTCTTAGAGCG	Used to sequence verify pEDH5
EH262	TGATCAATCTCAGATTTCAAATATAG	Annealed to EH263 for integration into pDB18
EH263	AAAACTATATTTGAAATCTGAGATTG	Annealed to EH262 for integration into pDB18
EH285	GCATCGTCTCATCGGTCTCATATGATGTTTGGCAAGAAGAAGAACAACGGAGGATCTTCC	Used with EH287 to amplify fragment 1 of *pol2-M644G* for cloning into pYTK001
EH287	GCATCGTCTCAAGATACTAGCGTTGCGTGCATGTTCG	Used with EH285 to amplify fragment 1 of *pol2-M644G* for cloning into pYTK001
EH288	GCATCGTCTCAATCTCAGGAAACCTTATCCAGTGG	Used with EH289 to amplify fragment 2 of *pol2-M644G* for cloning into pGEM-T Easy
EH289	GCATCGTCTCATCATCCCTCGCATGCTTATCAG	Used with EH288 to amplify fragment 2 of *pol2-M644G* for cloning into pGEM-T Easy

EH290	GCATCGTCTCAATGAGGCCAAGAAGA TGATTGTGC	Used with EH291 to amplify fragment 3 of *pol2-M644G* for cloning into pYTK001
EH291	GCATCGTCTCAGAGTTTCATATATATA GTTCAATGGGTC	Used with EH290 to amplify fragment 3 of *pol2-M644G* for cloning into pYTK001
EH292	GCATCGTCTCAACTCACAGTGAAAACA CGAT	Used with EH293 to amplify fragment 4 of *pol2-M644G* for cloning into pGEM-T Easy
EH293	GCATCGTCTCACACGACGCTTTCTATC TCTTG	Used with EH292 to amplify fragment 4 of *pol2-M644G* for cloning into pGEM-T Easy
EH294	GCATCGTCTCACGTGATCAATTATTTG GTAATACAAAC	Used with EH295 to amplify fragment 5 of *pol2-M644G* for cloning into pGEM-T Easy
EH295	ATGCCGTCTCAGGTCTCAGGATCCTAT GGTCAAATCAGCAATAC	Used with EH294 to amplify fragment 5 of *pol2-M644G* for cloning into pGEM-T Easy
EH301	GTCGGTTGCCAGTTCAATTG	Used to sequence verify pEDH11
EH302	AGTGCACTTGGCTCAACTATTG	Used to sequence verify pEDH11
EH303	CCGTCTTCATTGAAGACAGC	Used to sequence verify pEDH19
M13F	GTAAAACGACGGCCAG	Used to sequence verify fragments 2, 4 and 5 of *pol2-M644G* in pGEM-T Easy (pEDH9, 10, 11)
M13R	CAGGAAACAGCTATGAC	Used to sequence verify fragments 2, 4 and 5 of *pol2-M644G* in pGEM-T Easy (pEDH9, 10, 11)
MH012	CATGTGAGCAAAAGGCCAG	Targets BYP7434 plasmid Used with CA035 to genotype *URA3-pGAL-Os.TIR1* strains
MH022	GCAGATTGTACTGAGAGTGC	Used to sequence verify pEDH5
SB004	CTCCTCTCCAAGCCCATTT	Used to sequence verify pEDH9
SB007	TCCAGATGGTCGCTAAGCCTGTTAGAA GGAAGAGGCTGAGAACCCTAGATAAT TGGTACCGGCGTACGCTGCAGGTCGA C	Used with SB008 to amplify 3xminiAID sequence with *RNH201* 3' end homology arms for *RNH201* tagging with 3xminiAID
SB008	TGAAGTTATGACATATGTAGTATTACAT GAAGATATATAGTATGTGCAAACTGGA GGTGAGGACGAGGCAAGCTAAACAGA TCT	Used with SB007 to amplify 3xminiAID sequence with *RNH201* 3' end homology arms for *RNH201* tagging with 3xminiAID
SB011	TCCCGATGAGATCCTGGGTT	Targets *S. cerevisiae RNH201* Used with AC0019 to distinguish between the untagged and 3xminiAID-tagged genes

2.4 Yeast methods

2.4.1 Growth conditions

Yeast cells were frozen in 25% glycerol at -80 °C for long-term storage. To revive strains for experimental work, cells were streaked to single colonies on the appropriate medium from the glycerol stock and incubated at the optimal temperature according to species and strain. Once

grown, cells were stored at 4 °C on plates for up to two weeks. To set up starter liquid cultures, 5 mL of medium were inoculated with cells from a single colony and incubated in an air-incubated shaker overnight at 200 rpm. Experimental cultures were set up by diluting starter cultures to the desired cell density in the appropriate volume of medium and incubated in air- or water-incubated shakers. *S. cerevisiae* strains were grown at 30 °C, unless otherwise specified. *S. pombe* strains were also cultured at 30 °C, except the temperature sensitive YDP66 strain, which was grown at its permissive temperature of 25 °C and heat shocked at its restrictive temperature of 36 °C.

2.4.2 Measurement of growth rate

To calculate the doubling time of a strain, multiple optical density (OD) measurements were taken during exponential growth phase (at wavelength 595 nm for *S. pombe* and 600 nm for *S. cerevisiae*). Each measurement was relative to a cuvette containing the appropriate medium. An exponential trendline was fit to a plot of time vs OD measurements with the following equation:

$$y = Ae^{kt}$$

Where y = the OD reading at time t, A is the initial OD reading, and k is the growth rate. The doubling time of a strain was derived as follows:

$$Doubling\ time\ (\min) = \frac{\ln(2)}{k}$$

2.4.3 Measurement of cell number by haemocytometry

To determine the number of *S. cerevisiae* or *S. pombe* cells per mL of a culture, cells were counted using a haemocytometer. For this, a small aliquot of cells was taken from an exponentially growing culture, diluted if necessary, and briefly sonicated at 25 Watts using a Branson Digital Sonifier 250 to separate flocculated cells. The number of cells in at least five of the 25 central squares was counted and the mean number of cells per square was calculated. The number of cells per mL of culture was calculated using the following formula:

cells/mL = mean number of cells/square x 25 (number of squares) x dilution factor x 10^4

2.4.4 Producing diploids, sporulation, and tetrad dissection

Saccharomyces cerevisiae

Diploids were produced by either mating in liquid medium or on a plate. For liquid mating, a small amount of cells of each mating type was added to 200 µL of YPAD in a 1.5 mL tube and incubated at 30 °C, 200 rpm in an air-incubated shaker for the duration of a day. At the end of the incubation, 100 µL of the culture were spread on a KAc plate and incubated at room temperature until tetrads could be observed by microscopy. For the plate-based mating, one strain was patched onto a YPAD plate in a small area. The strain of the opposite mating type was then spread in this patch, so the two strains were mixed. Plates were incubated at 30 °C overnight. The following day, cells from the patch would be spread onto a KAc plate for sporulation and treated as above. Upon observation of tetrads, the ascus surrounding tetrads was digested by treatment with 1 µL of 5 mg/mL Zymolyase-20T in 19 µL of dissection buffer (0.8 M sorbitol, 10 mM ethylene dinitrilo tetraacetic acid (EDTA), 10 mM $NaPO_4$ pH 7.2) and incubation at 37 °C for 10-20 min, depending on the strain. The digested tetrads were diluted with 40 µL of Milli-Q H_2O, then approximately 20 µL were spread in a line down the centre of a YPAD plate. After drying, sister spores were separated on the plate using a Singer Instruments SporePlay dissection microscope. Plates of dissected tetrads were incubated at 30 °C until spores had produced colonies large enough for colony PCR and/or replica plating (typically around 3 days). Genotypes linked to a prototrophic or antibiotic-resistance gene were detected by replica plating tetrad dissection plates onto the appropriate medium and incubation at 30 °C overnight. Colonies with the prototrophic or antibiotic-resistance gene grew, while those lacking the marker did not, typically showing a 2:2 segregation within tetrads.

Schizosaccharomyces pombe

S. pombe diploids are usually unstable and sporulate immediately after mating between two haploids. To induce mating and sporulation, two strains of opposite mating type were mixed in a patch on a malt extract plate at room temperature until tetrads were observed (typically around 3 days). Sporulation was confirmed by observation of asci using a light microscope. Tetrad dissection was performed to isolate haploid offspring. Zymolyase pre-treatment of *S.*

pombe asci is not necessary, as the ascus wall breaks down naturally following incubation at 37 °C. Therefore, the tetrad dissection process began by isolating individual asci from the population on a YES plate. Plates were then incubated at 37 °C to allow the ascus wall to degrade. The isolated tetrads were then dissected and plates were incubated at 30 °C until colonies had grown from the individual spores (usually 3-4 days, depending on strains).

2.4.5 Yeast transformation

Saccharomyces cerevisiae

Cells were transformed as in (Gietz and Schiestl 2008). Briefly, an overnight pre-culture was diluted to early log phase allowing for 5 mL of cells per transformation. This culture was then left to grow to mid log phase (0.7-1.85x10^7 cells/mL). At this stage, cells were harvested by centrifugation at 1962 x *g* for 5 min. The cell pellet was rinsed twice with 25 mL of Milli-Q H_2O. The cells were then resuspended in 1 mL of Milli-Q H_2O, transferred to a 1.5 mL tube and centrifuged at 15,871 x *g* for 30 s. The supernatant was completely removed and the cell pellet was resuspended in a mix of: 240 µL of 50% (w/v) PEG3350, 36 µL of 1 M lithium acetate, 50 µL of single-stranded salmon sperm DNA, and 34 µL of DNA (in the range of 100 ng – 1 µg of plasmid DNA and/or 1 µg – 5 µg of double-stranded linear molecules). Resuspended cells were incubated in a 42 °C water bath for 40 min. Following the heat shock, tubes were centrifuged at 15,871 x *g* for 30 s and the supernatant was completely removed. For prototrophic gene selection, cell pellets were resuspended in Milli-Q H_2O and plated on the appropriate medium in ratios of 10% and 90% of cells per plate. Plates were incubated at 30 °C until colonies had grown. For antibiotic gene selection, cells were plated onto YPAD as above and incubated at 30 °C overnight before replica plating onto selective medium the following day. Alternatively, cells were resuspended in 1 mL of YPAD and incubated at 30 °C, 200 rpm for a minimum of 2 h before plating onto selective medium.

Schizosaccharomyces pombe

An overnight pre-culture was diluted to early log phase in EMM, low dextrose allowing for 50 mL of cells per transformation. The culture was left to grow to mid log phase (1x10^7 cells/mL)

before harvesting by centrifugation at 2,500 x *g* for 5 min. Cells were rinsed in 50 mL of Milli-Q H_2O, then resuspended in 1 mL of Milli-Q H_2O for transfer to 1.5 mL tubes. The cells were pelleted by centrifugation at 16,000 x *g* for 1 min and rinsed in 1 mL of LiAc-TE (0.1 M lithium actetate, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA). Following the rinse, cells were resuspended in LiAc-TE at $2x10^9$ cells/mL (250 µL per transformation). From this, 100 µL of cells were taken and mixed with 10 µL of 2 mg/mL single-stranded salmon sperm DNA and up to 10 µL of DNA (corresponding to up to 10 µg of linear DNA or 1 µg of plasmid DNA). The tubes were mixed gently and incubated at room temperature for 10 min before addition of 260 µL of a 40% (v/v) PEG-4000-LiAc mix. Tubes were incubated for 30-60 min at 30 °C without shaking. Following incubation at 30 °C, 43 µL of DMSO were added and tubes were mixed gently by inversion. Cells were heat shocked at 42 °C for 5 min, then pelleted by centrifugation at 6000 x *g* for 1 min. Cell pellets were rinsed in 1 mL of Milli-Q H_2O, then resuspended in 500 µL of Milli-Q H_2O and plated at 20% and 80% of cells per plate. For the hormone-binding domain plasmid transformation, YES plates were supplemented with 125 nM β-oestradiol to allow correct RNase H2B localisation.

2.4.6 Spot assay

Spot assays were used to detect growth defects in strains. Overnight pre-cultures were diluted in medium to obtain 100 µL of $4x10^6$ cells/mL. From this initial dilution, a 10-fold dilution series was made with a total of five dilutions per strain. Dilutions were kept on ice throughout to prohibit cell growth while 3 µL of each dilution were pipetted onto a plate of the indicated medium. Plates were dried and incubated at the indicated temperature before imaging.

2.4.7 Protein extraction

Total protein was extracted from *S. cerevisiae* for Western blot analysis, as previously described (Kushnirov 2000). Briefly, approximately 2.5 OD_{600} units of exponentially growing cells were harvested by centrifugation at 3488 x *g* for 2 min. Cell pellets were resuspended in 1 mL of Milli-Q H_2O, transferred to 1.5 mL tubes and centrifuged at 15,871 x *g* for 30 s. After completely removing any supernatant, cell pellets were frozen at -20 °C until extraction. For

alkali treatment, pellets were resuspended in 100 µL of Milli-Q H_2O, and 0.2 M NaOH was added to a final concentration of 0.1 M. Cells were incubated in this solution at room temperature for 5 min, then centrifuged at 15,871 x *g* for 30 s. The resulting cell pellets were resuspended in 50 µL of sodium dodecyl sulphate (SDS) sample buffer (60 mM Tris-HCl, pH 6.8, 5% glycerol, 2% SDS, 4% β-mercaptoethanol, 0.0025% bromophenol blue) and heated at 85 °C for 3 min. Debris was pelleted by centrifugation at 15,871 x *g* for 30 s and the protein-containing supernatant was transferred to a fresh 1.5 mL tube. Samples were loaded immediately on an SDS polyacrylamide gel (section 2.5.1). Remaining sample was stored at -20 °C.

2.4.8 Genomic DNA extraction

Quick DNA preparations

A quick DNA extraction method, without purification steps, was used to obtain the template for genotyping PCRs. A buffer of the following composition: 1 M sorbitol, 67.5 mM Na_2HPO_4, 15.8 mM $Na_2H_2PO_4$, pH 7.5 was supplemented with 2.5 mg/mL Zymolyase-100T and stored in 5 µL aliquots in 0.2 mL tubes at -20 °C. To release DNA, a small amount of cells from a colony was resuspended in this buffer and incubated in a PCR thermocycler using the following programme: 37 °C for 30 min to spheroplast the cells, 95 °C for 5 min to lyse the cells and denature enzymes, 15 °C “forever” to store the cell lysate until use. The samples were diluted with 50-100 µL of Milli-Q H_2O for use as PCR templates and were stored at 4 °C.

Depending on the downstream application, one of several DNA extraction protocols was used to isolate purified DNA from cells as described below.

Saccharomyces cerevisiae

i. For Okazaki fragment detection, labelling and repair trials, DNA was extracted according to (Smith, Yadav, and Whitehouse 2015).

In summary, 5 mL of SCE buffer (1 M sorbitol, 100 mM sodium citrate, 60 mM EDTA, pH 7.0) per 50 mL of cell culture were supplemented with 3.5 µL of β-mercaptoethanol before beginning each extraction. Cell pellets were rinsed twice in 1 mL of SCE, then resuspended in 900 µL of SCE. For each 50 mL of cells, 5 mg of Zymolyase-20T were resuspended in 100 µL of SCE, and 100 µL of this were added to each 900 µL of cells. The mix was incubated at room temperature for 3 min to break down the cell wall (spheroplast) the cells. The spheroplasts were collected by pulse centrifugation and rinsed twice in 1 mL of SCE. After the final SCE rinse, the spheroplast pellets were gently resuspended in 480 µL of lysis buffer (50 mM Tris-HCl pH 8.0, 50 mM EDTA, 100 mM NaCl, 1.5% w/v N-lauroylsarcosine) and incubated with 20 µL of 10 mg/mL proteinase K at 42 °C for 2 h. During the incubation, tubes were gently flicked periodically. If after 1.5 h of incubation, there was still visible cell debris, a further 5 µL of proteinase K were added and the tubes were incubated for 1 h. Following the degradation of proteins, 200 µL of 5 M potassium acetate were added to precipitate cell debris. The tubes were mixed by inversion and cooled on ice for at least 5 min before centrifugation at 16,000 x *g* for 20 min at 4 °C. The supernatant was transferred to a fresh 1.5 mL tube and 0.5 mL of isopropanol were added to precipitate the nucleic acids. Tubes were mixed by inversion and centrifuged at 16,000 x *g* for 10 min at 4 °C. The resulting nucleic acid pellets were rinsed in 0.5 mL of 70% (v/v) ethanol and allowed to dry before addition of 250 µL of STE (100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA). There was no active resuspension of the pellet. To degrade RNA, 1 µL of 20 mg/mL RNase A was added and the tubes were incubated at 37 °C for at least 1 h, with occasional gentle flicking to mix and facilitate pellet resuspension. Following digestion by RNase A, 30 µL of 3 M sodium acetate (pH 5.2) and 900 µL of absolute ethanol were added to precipitate the DNA. The tubes were inverted several times to mix before centrifugation at 5,000 x *g* for 10 min. The pelleted DNA was rinsed twice with 70% (v/v) ethanol and allowed to dry. TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) was added to each DNA pellet at 1 µL per mL of cell culture and the DNA was left to dissolve at 4 °C overnight. Once the DNA had dissolved, the tubes were centrifuged at 16,000 x *g* for 5 min and the supernatant was transferred to a fresh 1.5 mL tube.

ii. For pulsed-field gel electrophoresis or optical mapping by Irys or Saphyr, DNA was extracted by plug lysis to obtain ultra-long molecules according to instructions from BNG.

To harvest exponentially growing cells or cells that had undergone cell cycle synchronisation (section 2.4.9), cultures were transferred to 50 mL tubes and pelleted by centrifugation at 1,962 x *g* for 2 min at 4 °C. Each 50 mL cell pellet was rinsed in 5-10 mL of ice-cold 50 mM EDTA (pH 8.0, volume depended on culture OD_{600} a time of harvesting) and centrifuged at 1,962 x *g* for 2 min at 4 °C. Pellets were resuspended in 5-10 mL of ice-cold 50 mM EDTA (pH 8.0) and cells from the same culture were pooled into one tube. The cell density was assessed by counting the cell number using a haemocytometer (section 2.4.3) and the cells were pelleted by centrifugation. The cells were washed in 10 mL per 100 mL of culture of SE (1 M sorbitol, 25 mM EDTA pH 8.0) freshly supplemented with dithiothreitol (DTT) to a final concentration of 50 mM and pelleted as above. This was followed by a wash in 1 M ice-cold sorbitol at 10 mL per 100 mL of culture during which the cells were split into 10 mL aliquots and pelleted. Each cell pellet was gently resuspended in 10 mL of SCE buffer (1 M sorbitol, 1 mM EDTA, 10 mM sodium citrate, pH 5.8) and equilibrated at 30 °C for 10 min, before addition of 15 µL of 300 U/mL Zymolyase-100T, gentle swirling to mix and further incubation at 30 °C. Spheroplasting was assessed by observation under a light microscope after 5 µL of cells had been mixed with 5 µL of 5% SDS. Once spheroplasting was complete, 30 mL of ice-cold 1 M sorbitol were added and the tubes were gently mixed. Spheroplasts were collected by centrifugation at 1,962 x *g* for 3 min at 4 °C. Cell pellets were gently resuspended in cell suspension buffer (Bio-Rad: 10 mM Tris pH 7.2, 50 mM EDTA, 2 mM NaCl), the volume of which depended on the cell density as determined by OD_{600} measurements and cell counting. The volume varied between experiments to try to identify the optimal volume-to-cell-number value. For the final Saphyr samples, approximately 1 µL of cell suspension buffer was added per $5.7x10^6$ cells. Suspensions from the same sample were pooled in a 1.5 mL tube using a wide-bore pipette tip, then equilibrated at 43 °C for 4-10 min before use. Before the first use, a 12 mL bottle of 2% CleanCut Agarose (Bio-Rad, 170-3594) was melted and aliquoted into 500 µL aliquots in 1.5 mL tubes and stored at 4 °C. For each experiment, a 500 µL aliquot was incubated at 70

°C until completely melted, then equilibrated at 43 °C for at least 10 min before use. To each 150 µL of spheroplasts in cell suspension buffer, 90 µL of the 2% agarose were added (0.75% final agarose concentration) and the solution was mixed by pipetting up and down with a wide-bore tip with a P1000 set to 90% of the solution volume. Plugs were cast by adding the mix to each well of a 10-well plug mould (Bio-Rad, 170-3713) using a 200 µL wide-bore pipette tip, with mixing between each plug. Plugs were incubated at 4 °C until the agarose had solidified. Plugs were then digested with proteinase K to degrade cellular proteins. For this, 167 µL of Puregene Proteinase K (Qiagen, 158920) were mixed with 5 mL of IrysPrep lysis buffer (BNG) in a 50 mL tube and up to 5 plugs were incubated in the solution in a Thermomixer comfort (Eppendorf) at 50 °C for 2 h with intermittent mixing (10 s at 450 rpm followed by 10 min without mixing). A fresh proteinase K solution was prepared as above, and after 2 h, the plugs were transferred to the fresh solution and incubated overnight (alternatively, the overnight treatment was performed first, followed by the 2 h treatment the next day). Next, 50 µL of RNase A (Qiagen, 158924) were added to each tube and the samples were incubated at 37 °C with intermittent mixing as above, for at least 1 h. Plugs were then rinsed three times with 10 mL of wash buffer (20 mM Tris pH 8.0, 50 mM EDTA) by gently swirling the plugs in the wash buffer for around 10 s and discarding the solution. After the three quick rinses, plugs were rinsed four times in 10 mL of wash buffer for 15 min each rinse on a horizontal platform mixer or in a Thermomixer comfort (Eppendorf) at room temperature, 300 rpm. Plugs could be stored in the final wash buffer rinse at 4 °C for up to two weeks before proceeding. The rinsing was repeated five times using 10mL of TE (10 mM Tris, 1 mM EDTA pH 8.0) per 50 mL tube. The TE of the final rinse was discarded and each plug was transferred to an individual 1.5 mL tube. The tubes were pulse centrifuged to collect the plug at the bottom of the tube and the agarose was melted by incubation at 70 °C for 2-3 min followed by transfer to 43 °C for 5 min. To each tube, 2 µL of 0.5 U/µL Agarase (Thermo Scientific, EO0461) were added and stirred in with a pipette tip for 10 s. The tubes were incubated at 43 °C for 45 min. To clean the DNA solution, drop dialysis was performed using a 0.1 µm nitrocellulose membrane (Merck Millipore, VCWP04700) on TE for 45 min at room temperature, then the DNA was transferred to a fresh 1.5 mL tube. The DNA was left to equilibrate at room temperature overnight. The DNA

concentration was assessed the following day using Qubit dsDNA BR Assay Kit reagents (Invitrogen, Q32850). For accurate pipetting of the high molecular weight DNA for measurement of concentration, a positive-displacement pipette was used (Gilson, MICROMAN, F148501).

Schizosaccharomyces pombe

i. DNA was isolated from cells for agarose gel electrophoresis and Southern blotting as follows:

Mid-exponential phase cells (OD_{595} ≈ 0.5) were harvested by centrifugation at 4,815 x *g* at 4 °C for 5 min. Cells were rinsed in Milli-Q H_2O, and the same samples were pooled into a single 50 mL tube before centrifugation as before. The following volumes were used for each 200 mL of original cell culture. Cell pellets were resuspended in 500 µL of nuclear isolation buffer (NIB: 17% glycerol, 50 mM 3-(N-morpholino)propanesulphonic acid (MOPS), 150 mM potassium acetate, 40 mM EDTA, pH to 7.2 with KOH, filter-sterilised through a 0.22 µm pore filter) with 5 mg/mL Zymolyase-20T and incubated at 37 °C for approximately 30 min. Spheroplasting progression was monitored by mixing 1 µL of the cells with 9 µL of 0.5% (w/v) SDS and observation under a microscope. Once the majority of cells had been spheroplasted, 5 mL of ice cold Milli-Q H_2O were added and the samples were mixed by vortexing. Cells were pelleted by centrifugation as above and the pellets were gently resuspended in 0.5 mL of QIAGEN buffer G2 (800 mM guanidine hydrochloride, 30 mM Tris-Cl, pH 8.0, 30 mM EDTA, pH 8.0, 5% Tween-20, 0.5% Triton X-100). For RNA degradation, 25 µL of 10 mg/mL RNase A were added and the samples incubated at 37 °C for at least 30 min. Next, 25 µL of 30% N-lauroyl sarcosine and 25 µL of 20 mg/mL proteinase K were added and the samples incubated at 55 °C for at least 1 h to degrade proteins. Samples were centrifuged at 4,815 x *g* at 4 °C for 15 min and the supernatant was transferred to a 1.5 mL tube. The pellet was gently resuspended in 250 µL of QIAGEN buffer G2 and 12.5 µL each of 30% N-lauroyl sarcosine and 20 mg/mL proteinase K were added before incubation at 55 °C overnight. The next day, this tube was centrifuged at 4,815 x *g* at 4 °C for 15 min and the supernatant was combined with the first

supernatant. A QIAGEN 100/G Genomic-tip was equilibrated with 4 mL of QIAGEN buffer QBT (750 mM NaCl, 50 mM MOPS, pH 7.0, 15% isopropanol, 0.15% Triton X-100). To purify genomic DNA from the cell lysate, the supernatant was applied to the equilibrated Genomic-tip and processed according to manufacturer's instructions. In brief, after allowing the liquid to flow through the column, the column-bound DNA was rinsed twice with 7.5 mL of QIAGEN buffer QC (1 M NaCl, 50 mM MOPS, pH 7.0, 15% isopropanol), then eluted in 5 mL of QIAGEN buffer QF (1.25 M NaCl, 50 mM Tris-Cl, pH 8.5, 15% isopropanol). Once the eluate was collected in a fresh 50 mL tube, DNA was precipitated by addition of 3.5 mL of isopropanol and centrifugation at 4,815 x *g* at 4 °C for at least 1 h. The resulting pellet was rinsed in 70% (v/v) ethanol and centrifuged at 4,815 x *g* at 4 °C for 10 min. Once the pellet had air-dried, 50 µL of TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) were added and the DNA allowed to resuspend at 4 °C overnight, or at room temperature.

ii. Initially, DNA was extracted for BrdU-IP and nanopore sequencing as published in (Murray, Watson, and Carr 2016), with slight modification. Later experiments used the protocol described in point iii.

Cells were pelleted by centrifugation at 1,962 x *g* for 5 min. Each cell pellet was resuspended in 1 mL of Milli-Q H_2O and split into two 1.5 mL tubes (for 40 mL of cells harvested). The cells were pelleted by centrifugation at 17,000 x *g* for 1 min. Cell pellets were stored at -20 °C until DNA extraction. For 20 mL cell volumes, the cell pellets were resuspended in CSE buffer (1.2 M sorbitol, 40 mM EDTA (pH 8.0) in 50 mM citrate-phosphate buffer (pH 5.6)) to a total volume of 1 mL. The cells were split into two 1.5 mL tubes and 500 µL of CSE buffer containing 2 mg/mL Zymolyase 20T were added to each tube. Tubes were incubated in a 37 °C water bath for 15 min and cell wall digestion was checked by combining 5 µL of cells with 5 µL of 10% SDS and observation under the microscope. Cells underwent further incubation at 37 °C if necessary. Once the spheroplasting was complete (when > 95% of the cells had been spheroplasted) samples were centrifuged at 17,000 x *g* for 1 min and resuspended in 450 µL of 5x TE (50 mM Tris-Cl (pH 7.5), 5 mM EDTA (pH 8.0)). To lyse the cells, 50 µL of 10% SDS

were added and tubes were mixed by inversion and incubated for 5 min at room temperature. To precipitate cell debris, 150 µL of 5 M potassium acetate were added and tubes were mixed by inversion and incubated on ice for 10 min. The cell lysate was centrifuged at 17,000 x *g* for 10 min and the supernatant was transferred to a 1.5 mL screw cap tube containing 600 µL of isopropanol to precipitate the nucleic acid. Tubes were mixed by inversion and centrifuged at 17,000 x *g* for 10 min. The resulting pellet of nucleic acid was washed with 75% (v/v) ethanol and resuspended in 500 µL of 5x TE supplemented with 5 µL of 10mg/mL RNase A per 495 µL. Tubes were incubated at 37 °C for 15-75 min. To degrade residual protein, 4 µL of 10% SDS, 10 µL of 1x TE, and 10 µL of 20 mg/mL proteinase K were added to each tube, followed by incubation at 55 °C for 60-75 min. Samples were allowed to cool to room temperature before addition of 500 µL of phenol:chloroform:isoamyl alcohol (50:49:1) and mixing by inversion. Phases were separated by centrifugation at 17,000 x *g* for 5 min. The upper aqueous phase was transferred to a fresh 1.5 mL screw cap tube and a further 500 µL of phenol:chloroform:isoamyl alcohol (50:49:1) were added. The tubes were mixed by inversion before centrifugation at 17,000 x *g* for 5 min. The upper aqueous phase was transferred to a fresh 1.5 mL screw cap tube and 500 µL of chloroform was added to remove residual phenol. The tubes were mixed by inversion before centrifugation at 17,000 x *g* for 5 min. The upper aqueous phase was transferred to a new 1.5 mL tube and the DNA was precipitated by addition of 25 µL of 3 M sodium acetate (pH 5.2) and 500 µL of isopropanol before incubation for 10 min on ice. The DNA was pelleted by centrifugation at 17,000 x *g* for 10 min and then rinsed in 75% (v/v) ethanol. Each DNA pellet was resuspended in 100 µL of 1x TE. The purity of the extracted DNA was assessed by measurement of A260/A230 and A260/A280 using a DeNovix Spectrophotometer. Where samples had been split into multiple tubes, the corresponding samples were pooled and the concentration was determined using a Qubit dsDNA HS Assay kit.

iii. To obtain high molecular weight DNA for nanopore sequencing, DNA was isolated as follows.

Cells were pelleted by centrifugation at 1,962 x *g* for 5 min. They were rinsed in Milli-Q H_2O, then transferred to 2 mL tubes at an equivalent of 20 mL of cell culture per tube. Cells were pelleted again by centrifugation at 5,000 x *g* for 1 min and the supernatant was completely removed. Cell pellets were resuspended in 1 mL of solution 1 (1.2 M sorbitol, 200 mM Tris-HCl, pH 7.5, 20 mM EDTA) supplemented with β-mercaptoethanol to a final concentration of 1% and 1 mg of zymolyse-20T. The pellet was resuspended by pipetting up and down. Cell wall digestion was carried out at 37 °C for at least 30 min and spheroplasts were pelleted by a short pulse of centrifugation. For each cell pellet, 500 µL of solution 2 (50 mM Tris-HCl, pH 7.5, 100 mM sodium chloride, 100 mM EDTA, 0.5% (w/v) SDS) were supplemented with sorbitol to a final concentration of 89 mM and 100 µg of RNase A, and 520 µL of this mix were added to each tube. The tubes were gently inverted several times to mix and incubated at room temperature for at least 1 h. Between 300-600 µg of proteinase K were added to each tube and tubes were again mixed by inversion and incubated at 65 °C for between 2 and 6 h with occasional mixing by inversion. The final amount of proteinase K added to each tube varied within the indicated range between experiments to try and find the optimal amount for complete removal of observable insoluble protein. DNA was extracted from the mix by addition of 550 µL of phenol:chloroform:isoamyl alcohol (50:49:1) to each tube and incubation on a tube rotator for 15-30 min. Centrifugation at 9,391 x *g* for 2 min was used to separate the organic and aqueous phases allowing the upper aqueous phase to be slowly and carefully transferred to a clean 1.5 mL tube using a wide-bore pipette tip. The phenol:chloroform:isoamyl alcohol steps were repeated. To remove residual phenol from the aqueous phase, 900 µL of chloroform were added to each tube and the tubes were processed as above. The chloroform steps were repeated until the phenol smell was no longer detectable. To precipitate the DNA, 1 mL of absolute ethanol were added to each tube and the tubes were inverted to mix. If the DNA formed a white spool, this was carefully transferred to a fresh tube containing 70% (v/v) ethanol to rinse the DNA. The ethanol was removed and the DNA was rinsed a second time

to remove residual salts. In some cases, the DNA did not spool, instead it separated into an aqueous phase at the bottom of the tube. When this happened, the supernatant was carefully removed without disturbing the DNA, and the pellet was rinsed twice with 70% (v/v) ethanol. After the ethanol rinses, the DNA was dried by incubation at 50 °C with open lids for several minutes. The resulting DNA was resuspended in 0.1x TE (1 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) by incubation at 50 °C for at least 20 min, or overnight at 4 °C.

iv. For Irys analysis, wild-type *S. pombe* DNA was prepared as described above for *S. cerevisiae* (ii) with the following differences:

To harvest the cells, an exponentially growing culture was transferred to 50 mL tubes and centrifuged at 1,962 x *g* for 5-10 min. Pellets were resuspended in 5 mL of CSE buffer (20 mM citrate/phosphate buffer, pH 5.6, 1.2 M sorbitol, 40 mM EDTA) and cells from the same culture were pooled. Cells were pelleted by centrifugation as above and rinsed a second time in CSE buffer. To spheroplast the cells, the pellet was resuspended in 5 mL of CSE buffer per 100 mL of cell culture supplemented with 1 mg/mL Zymolyase 20-T and incubated at 37 °C until completion (at least 30 min). Spheroplasts were pelleted by centrifugation at 1,962 x *g* for 10 min, then resuspended in cell suspension buffer (Bio-Rad: 10 mM Tris pH 7.2, 50 mM EDTA, 2 mM NaCl). Subsequent steps were identical to those described for *S. cerevisae* above.

2.4.9 Cell cycle synchronisation

Saccharomyces cerevisiae

The YK402 and TKY375 strains both have the *bar1Δ* genotype, which means that they do not produce the aspartyl protease, Bar1, to degrade the mating pheromone, α-factor. This enabled cells to be arrested in G1 phase by one addition of α-factor to the culture. Cells were grown in YPA + 2% raffinose (YPR) at 30 °C, 200 rpm in either air- or water-incubated shakers until they reached an OD_{600} of approximately 0.2. A sample was taken from the asynchronous culture for analysis of total DNA content by flow cytometry. To synchronise cells in G1, α-factor was added to a final concentration of 200 nM and cultures were incubated for 2 h. The arrest

was monitored by microscopy. For induction of *TIR1* expression (in TKY375), a protein required for the auxin-inducible degron (AID) system, galactose was added to a final concentration of 2% and cultures were incubated for a further 1 h. Indole-3-acetic acid (IAA) was added to a final concentration of 500 µM to cell cultures 1 h after galactose addition. Cultures were incubated for 1 h in the presence of IAA to allow for the destruction of AID-tagged DNA ligase in TKY375 before entry into S phase. Pronase protease was added to a final concentration of 2 µg/mL to cultures to degrade α-factor, thereby allowing release into S phase. Samples for DNA extraction and flow cytometry analysis of total DNA content were taken at the indicated time points.

Schizosaccharomyces pombe

The temperature-sensitive *cdc10-m17* strain grows at the permissive temperature of 25 °C but is arrested in G1 phase of the cell cycle when incubated at 35 °C (Nurse, Thuriaux, and Nasmyth 1976). For cell cycle synchronisation experiments, cells were grown in YES to OD_{595} of 0.2 at 25 °C, 200 rpm in either air- or water-incubated shakers. A sample was taken from the asynchronous culture for analysis of total DNA content by flow cytometry. The culture was shifted to 36 °C in a water-incubated shaker (200 rpm) for approximately one doubling time (~3.5 – 3.67 h). The arrest progression was monitored by taking samples at the indicated time points for flow cytometry. For BrdU experiments, the indicated concentration of BrdU was added to the culture 3 h and 15 min after the shift to 36 °C. Cells were released into S phase from the G1 phase arrest 3 h and 35 min after the shift to 36 °C by rapidly cooling the culture in a bath of icy water until the temperature of the culture reached 25 °C. The cells were then placed in a 25 °C water-incubated shaker (200 rpm) and S phase progression was monitored by taking samples for flow cytometry at the indicated time points.

In experiments where the G1 arrest was followed by release into hydroxyurea (HU), then subsequent release into S phase, HU was added to a final concentration of 20 mM 20 min prior to the release from the temperature arrest. Once returned to 25 °C as described above, the cultures were incubated for 1 h in the presence of HU. To chase the BrdU following the

release from HU, thymidine was added to 10x the concentration of BrdU during the HU arrest, 45 min after the shift to 25 °C (15 min prior to the release from HU). To remove HU from cultures, thereby allowing S phase progression, it was necessary to wash the cells into fresh medium. This was achieved by applying the culture to a pre-wetted 150 mL filter and allowing the medium to flow through under the influence of an applied vacuum. The cells were rinsed by removing the vacuum, applying some fresh medium, then re-applying the vacuum to allow the medium to flow through. This was done a total of three times to ensure maximal removal of HU. After the final wash, the vacuum was removed and 25-50 mL of fresh, pre-warmed YES medium were added to the cells on the membrane. The cells were resuspended by aspirating and expelling the medium in a 25 mL pipette. The resuspended cells were transferred to a clean flask containing the required extra volume of medium, supplemented with thymidine at 10x the concentration of BrdU that was added during the G1 arrest and returned to a water-incubated shaker at 25 °C, 200 rpm for the remainder of the experiment.

For the multiple BrdU additions experiment (section 3.3.2), a starting 400 mL culture in a 2 L flask was arrested in G1 phase as above. Flow cytometry samples were harvested throughout the experiment to monitor the cell cycle synchronisation. After 3 h 15 min in the G1 arrest, 40 mL of cells were harvested for DNA extraction as a control sample. Then, 95 mL of the culture were transferred to a separate 500 mL flask and BrdU was added to a final concentration of 0.5 µM (“BrdU at -20 min” sample). After a further 20 min at the restrictive temperature (3 h 35 min in total), another 40 mL of cells were harvested from the culture to which BrdU had been added at 3 h 15 min, and both flasks (± BrdU) were cooled to 25 °C for release from G1 phase in parallel. The remaining culture in the 2 L flask was split equally between five 500 mL flasks at 25 °C (approximately 45 mL of culture per flask). To one of these aliquots, BrdU was added to a final concentration of 0.5 µM immediately following the temperature shift and culture splitting (“BrdU at 0 min”). To the remaining four aliquots, which had not yet been exposed to BrdU, 0.5 µM BrdU was added to each at 15 min intervals, with the final addition at 60 min after release from G1 (“BrdU at 60 min”). Thus, a total of six separate cultures from one original

starting culture had BrdU added at different time points during the cell cycle synchronisation. All remaining cells were harvested for DNA extraction 90 min after release from the G1 arrest.

2.4.10 Measurement of DNA content by flow cytometry

Saccharomyces cerevisiae

Sodium azide was added to samples to a final concentration of 0.1% to cause cell death by inhibition of respiration. The harvested cells were kept on ice until all samples had been collected. Cells were then pelleted by centrifugation at 15,871 x *g* for 30 s. Cell pellets were rinsed in 1 mL of Milli-Q H_2O and centrifuged again. Ethanol fixation was carried out by addition of 1 mL of 70% (v/v) ethanol to cell pellets, vortexing and incubation at room temperature for at least one hour, or overnight at 4 °C. Following ethanol fixation, cells were rinsed twice in 1 mL of 50 mM sodium citrate (pH 7.0). All remaining centrifugation steps were at 15,871 x *g* for 1 min. After the second rinse in sodium citrate, cells were treated with RNase A by resuspending cell pellets in 500 µL of 50 mM sodium citrate containing 0.25 mg/mL RNase A and incubation at 50 °C for at least 1 h. Proteinase K was then added to a final concentration of 769 µg/mL and samples were incubated at 50 °C for at least 1 h. Following proteinase K treatment, cells were pelleted and the supernatant completely removed, before resuspension in 500 µL of 1 µM SYTOX Green Nucleic Acid Stain (ThermoFisher Scientific, S7020), 0.1% sodium azide in 50 mM sodium citrate. SYTOX Green Nucleic Acid Stain fluorescence increases upon binding to nucleic acids, enabling the determination of cellular DNA content and thus inference of the cell cycle stage. DNA staining was carried out for at least 1 h at room temperature or overnight at 4 °C in light-protected conditions. Before flow cytometric analysis, cell clumps and doublets were disrupted by sonication using a Branson Digital Sonifier 250 with a tapered microtip set at 10% amplitude for two 3 s pulses with a 2 s interval. Samples were analysed using a BD LSRFortessa X-20 cell analyser with the 488 nm laser and 530/30 filter. The PMT voltages used were as follows: FSC = 610 V, SSC = 269 V, 488-530/30 = 450 V. For each sample, 100,000 events were recorded. Gating for single cells was subsequently carried out using FlowJo software and the distribution of event numbers against DNA content were plotted as histograms to observe cell cycle stage.

Schizosaccharomyces pombe

Cells were processed as described for *S. cerevisiae* with the following exceptions. The initial centrifugation steps up to the ethanol fixation were carried out at 2,700 x *g* at 4 °C for 5 min, with subsequent centrifugations at 3,000 x *g* for 5 min (room temperature). The sonication settings were: 20% amplitude for 5 s in 1 s pulses with 0.3 s intervals. The PMT voltages for analysis on a BD LSRFortessa X-20 cell analyser with the 488 nm laser and 530/30 filter were as follows: FSC = 505 V, SSC = 200 V, 488-530/30 = 423 V.

G1 phase cells that had undergone cytokinesis during the cell cycle arrest at the restrictive temperature had a DNA content of 1 genome copy (1C), while G2 phase cells had a DNA content of 2C, with S phase cells on a continuous scale between 1-2C. This analysis enabled the determination of the optimal conditions for cell cycle synchronisation experiments and verification of each cell cycle arrest and release.

2.5 Biochemical techniques

2.5.1 SDS polyacrylamide gel electrophoresis (PAGE)

Proteins were separated by size by SDS-PAGE. SDS-PAGE gels were made according to the following recipe (for a 7% gel):

Reagent	**Resolving gel**	**Stacking gel**
Milli-Q H_2O	4.97 mL	2.25 mL
Tris (1.5 M, pH 8.8 for resolving/0.5 M, pH 6.8 for stacking)	2.5 mL	1 mL
30% (w/v) acrylamide	2.33 mL	670 µL
10% (w/v) SDS	100 µL	40 µL
10% (w/v) ammonium persulphate (APS)	100 µL	40 µL
TEMED	4 µL	4 µL

The resolving gel was made first, poured into a plastic cassette (8 cm x 8 cm x 1 mm), a layer of isopropanol was added and the gel was allowed to set. Once the resolving gel had set, the isopropanol was poured off and the top edge of the gel was rinsed with distilled H_2O and dried

with Whatmann paper, before pouring on the stacking gel. A comb was added and the gel allowed to set. Once the gel had set, the cassette was placed in the electrophoresis tank and 1x running buffer (192 mM glycine, 25 mM Tris, 0.1% (w/v) SDS) was added to fill the tank. Removal of the comb allowed sample and ladder loading. PageRuler Prestained Protein Ladder (Thermo Scientific) was used to indicate protein sizes. Following protein extraction (section 2.4.7), 10 µL of each sample were loaded. Protein gels were run using the Invitrogen XCell SureLock Mini-Cell Electrophoresis system at 100-120 V until the dye front had reached the end of the gel. Gels were removed from the cassette and rinsed in distilled H_2O prior to processing for Western blot.

2.5.2 Western blot

Polyvinylidene difluoride (PVDF) membrane was used for Western blotting of proteins from SDS-PAGE gels. The membrane was cut to size and activated for 30 s in methanol before equilibration in transfer buffer (1x running buffer + 20% (v/v) methanol) for several minutes. SDS-PAGE gels, filter papers and fibre pads were also equilibrated in transfer buffer for several minutes. Transfers were performed using a Bio-Rad Mini Trans-Blot Electrophoretic Transfer Cell according to manufacturer's instructions. Briefly, after pre-soaking the gel, membrane and two filter papers per transfer, the gel sandwich was prepared in a cassette by layering a fibre pad on the black surface of the cassette, followed by a filter paper, the gel, the membrane, a filter paper, and the second fibre pad. Bubbles were removed by gently rolling across the second filter paper. The cassette was closed and placed into the blot module. A frozen cool block was added to the tank before filling with transfer buffer. Transfers were carried out at 4 °C at 30 V overnight. Following the transfer, the membrane was rinsed thoroughly in Milli-Q H_2O before staining in Ponceau Red solution (0.1% (w/v) Ponceau Red, 5% (v/v) acetic acid) to assess total protein transfer. Ponceau Red stain was removed by rinsing in phosphate-buffered saline (PBS) supplemented with 0.1% (v/v) Tween-20 (PBST) for approximately 30 min. The membrane was incubated in blocking buffer (3% (w/v) non-fat milk powder in PBST) on a 2D horizontal rocker for 1 h at room temperature. FLAG-tagged proteins were probed for using Monoclonal ANTI-FLAG M2 antibody produced in mouse

(Sigma, F1804). This was diluted to a final concentration of 1 µg/mL in 3% blocking buffer in a 50 mL tube to which the membrane was added. The probing was carried out on a 50 mL tube roller at room temperature for 1 h. Subsequently, the membrane was rinsed four times in 1% blocking buffer, each for 5 min on a 2D horizontal rocker at room temperature. The secondary antibody, Peroxidase AffiniPure Goat Anti-Mouse IgG (H+L, Jackson ImmunoResearch, 115-035-062), was diluted to a final concentration of 100 ng/mL in 1% blocking buffer. The rinsed membrane was incubated in the secondary antibody solution on a 2D horizontal rocker at room temperature for 1 h. The membrane was then rinsed four times in PBST, each for 5 min on a 2D horizontal rocker at room temperature. The horse-radish peroxidase (HRP) conjugated to the secondary antibody catalyses the reaction between luminol and peroxide, resulting in the emission of light, thereby enabling detection of the protein of interest. Thus, following the membrane rinses, Amersham ECL Prime Western Blotting Detection Reagent was used, according to manufacturer's instructions. Briefly, equal volumes of Solutions A and B were mixed. The membrane was removed from PBST and any excess liquid was drained before placing the membrane on a sheet of Saran wrap. The Solution A+B mix was applied to the membrane and incubated for 1 min at room temperature. Excess solution was then drained off the membrane before placing between two clean plastic sheets for imaging using an ImageQuant LAS 4000 biomolecular imager (GE Healthcare). The light emitted from the HRP reaction was detected by the chemiluminescence setting, while the protein ladder was imaged with white light to allow comparison of the antibody-bound protein with the sizes of the known standards.

2.6 Molecular genetic techniques

2.6.1 Polymerase chain reaction (PCR)

PCR was used to amplify regions of interest or for genotyping strains. BIOTAQ DNA polymerase (Bioline, BIO-21060) was generally used, unless high fidelity amplification of repair constructs or cloning fragments was required, in which case Phusion High-Fidelity DNA polymerase (NEB, M0530) was used. Master mixes of reagents were prepared when multiple reactions were performed by combining 1x NH_4 reaction buffer (Bioline), 2 mM magnesium

chloride (Bioline), 200 µM of each dNTP, 1 µM of each oligonucleotide primer, 0.5 units of BIOTAQ polymerase, and 0.5-1 µL of template DNA from a quick preparation (section 2.4.8) per 10 µL of reaction mix. PCRs were generally performed in 10 µL reaction volumes in 0.2 mL thin-wall PCR tubes. Reactions were carried out in a Bio-Rad DNA engine Tetrad 2 Peltier thermal cycler. Typically, PCR programmes began with a 95 °C incubation for 5 min for initial denaturation, followed by repeated cycles of: 95 °C for 30 s, primer annealing at the optimal temperature (usually in the range of 50-60 °C) for 30 s, primer extension at 72 °C for a duration dependent on the target fragment length, allowing 15-30 s per kb. Phusion High-Fidelity reactions were carried out in 1x GC buffer (NEB), with 50 µM of each dNTP, 0.5 µM of each oligonucleotide primer, approximately 1-3 ng of template DNA, and 0.1 units of Phusion polymerase per 10 µL reaction volume. For this enzyme, PCR programmes began with an initial denaturation step at 98 °C for 1 min, and PCR cycles generally comprised: 98 °C for 10 s, primer annealing at the optimal temperature (usually in the range of 50 – 72 °C) for 30 s, primer extension at 72 °C for a duration dependent on the target fragment length, allowing 15-30 s per kb. The cycling was followed by a final extension incubation at 72 °C for 5 min.

PCR amplicons were purified from other reaction components where necessary, using a GenElute PCR Clean-Up kit (Sigma, NA1020) according to manufacturer's instructions.

2.6.2 Cloning methods

Standard cloning

To produce the *rnh202-HBD* plasmid, pEDH5, standard cloning procedures were used. In brief, a PCR product of the desired insert (the *S. pombe rnh202*$^{+}$ gene with flanking restriction enzyme cut sites) was produced and purified from other reaction components using a GenElute PCR Clean-Up kit. The amplicon concentration was assessed spectrophotometrically and 150 ng of the product were digested with 5 units of *Pac*I (NEB, R0547) and 10 units of *Sal*I-HF (NEB, R3138) in 1x CutSmart buffer (NEB, B7204) in a final volume of 10 µL at 37 °C for 1 h. In a parallel reaction, 150 ng of the p637 vector were also digested with the same enzymes. After the digestion, 40 µL of Milli-Q H_2O were added to each

tube and both samples were purified from other reaction components using a GenElute PCR Clean-Up kit. For ligation of the digested PCR product into the digested plasmid backbone, 50 ng of each were combined with 1x T4 DNA ligase buffer (NEB, B0202), 200 units of T4 DNA ligase (NEB, M0202) and Milli-Q H_2O to a final volume of 10 µL. The reaction was incubated at room temperature overnight before transformation into chemically competent *E. coli*.

PCR A-tailing

For cloning into pGEM-T Easy, PCR amplicons were first A-tailed according to a protocol from NEB. Each PCR was performed in a final volume of 30 µL and the total product was purified from other reaction components using a GenElute PCR Clean-Up kit according to manufacturer's instructions. The final volume of the purified amplicons was reduced to 10 µL each using a Genevac miVAC DNA Vacuum-Integrated Centrifugal Concentrator System. Then, for each PCR, a 50 µL A-tailing reaction was set-up (a master mix excluding the PCR DNA was made when multiple PCRs were A-tailed in parallel): 1x ThermoPol Buffer (NEB, B9004), 0.2 mM dATP, 1 unit of *Taq* DNA polymerase (NEB, M0267), Milli-Q H_2O to make each reaction to 50 µL final volume. The total volume of the purified PCR amplicons was added to each reaction. These reactions were set up in 0.2 mL thin-wall PCR tubes and incubated in a Bio-Rad Tetrad thermal cycler at 72 °C for 20 min. A-tailed DNA was then purified from other reaction components using a GenElute PCR Clean-Up kit according to manufacturer's instructions.

Cloning into pGEM-T Easy vector

Cloning with the pGEM-T Easy vector was performed as described by the manufacturer in half the standard reaction volume. Briefly, 25 ng of pGEM-T Easy vector (Promega, A137A) were used for each reaction. The PCR product was added in a 1:1 molar ratio to the vector. The amount in ng of PCR product to add was calculated for each amplicon using the following formula:

$$ng_{insert} = \frac{ng_{vector} \times length\ (bp)_{insert}}{length\ (bp)_{vector}}$$

where ng_{insert} represents the calculated ng of insert PCR DNA required, ng_{vector} represents the amount in ng of pGEM-T Easy vector (in this case 25 ng). The length of the pGEM-T Easy vector was 3,015 bp. Each reaction included 25 ng of the pGEM-T Easy vector, 1x rapid ligation buffer (Promega, C671A), 1.5 units of T4 DNA ligase (Promega, M180A), Milli-Q H_2O to make the volume to 5 µL. Reactions were incubated at 4 °C overnight.

Bacterial transformation

Plasmids were transformed into competent DH5α *E. coli* cells (made in-house by Dr. Dzmitry Batrakou) by addition of the ligated plasmid to a 100 µL aliquot of cells and incubation on ice for 30 min. The cells were then heat shocked at 42 °C for 30 s and incubated on ice for 3 min. LB medium was added to each aliquot of cells (100 µL of medium per 100 µL of cells) and samples were incubated at 37 °C, 200 rpm for 1 h to allow cells to recover and synthesise the antibiotic resistance protein conferred by the plasmid. After recovery, cells were plated onto LB medium supplemented with the appropriate antibiotic ± ChromoMAX IPTG/X-Gal and incubated at 37 °C overnight.

For transformation with pGEM-T Easy-based plasmids, 2.5 µL of each 5 µL ligation reaction were added to a 100 µL aliquot of cells. Cells were selected for on ampicillin and with blue/white selection conferred by IPTG/X-Gal.

Isolating plasmids from *E. coli*

To test bacterial colonies following transformation with a plasmid of interest, 3-5 mL of LB medium supplemented with the appropriate antibiotic were inoculated with cells from a colony and incubated at 37 °C, 200 rpm overnight in an air-incubated shaker. Once cultures had reached stationary phase, cells were pelleted by centrifugation at 4,415 x *g* for 7 min. Each cell pellet was rinsed in 10 mL of PBS and centrifuged as before. After complete removal of the supernatant, plasmid DNA was isolated from cells using a NucleoSpin Plasmid kit (Macherey-Nagel, 740588.50), according to the manufacturer's instructions.

Golden Gate assembly

The Golden Gate assembly of a cassette plasmid encoding *pGAL-pol2-M644G-3xFLAG-6xHis-tTDH1* using the Yeast Toolkit system (Lee et al. 2015) required two sequential rounds of cloning (Figure 4.18).

The first round involved the production of a *pol2-M644G* part plasmid for use in the cassette plasmid assembly. The *pol2-M644G* gene was amplified as five fragments, the largest three of which were cloned into the pGEM-T Easy vector for sequence verification by Sanger sequencing (pEDH9, pEDH10, pEDH11). Part plasmid assembly of the five fragments into the pYTK001 entry vector was performed as described in (Lee et al. 2015). In summary, each Golden Gate assembly reaction comprised: 20 fmol of each DNA insert and plasmid, 1x T4 DNA ligase reaction buffer (NEB, B0202), 1,500 units of T7 DNA ligase (NEB, M0318), 5 units of *Bsm*BI (NEB, R0580), Milli-Q H_2O to a final volume of 10 µL. Reactions were incubated in a Bio-Rad Tetrad thermal cycler with the following programme: digestion by *Bsm*BI at 42 °C for 2 min, ligation by T7 DNA ligase at 16 °C for 5 min, repeated 24 times (total of 25 cycles), final digestion at 60 °C for 10 min, enzyme inactivation at 80 °C for 10 min. Following this assembly, 2.5 µL of each reaction were used to transform a 100 µL aliquot of DH5α cells as describe above. The pYTK001 entry plasmid conferred chloramphenicol resistance to successful transformants, so after recovery in LB medium, cells were plated onto LB supplemented with chloramphenicol.

The second round of cloning aimed to create the *pGAL-pol2-M644G-3xFLAG-6xHis-tTDH1* cassette plasmid. Cassette plasmid assembly involved *Bsa*I digestion, rather than *Bsm*BI. The reaction consisted of: 20 fmol each of pYTK096 (a pre-assembled *URA3* integration vector), pYTK030 (type 2 *pGAL1* plasmid), pEDH19 (type 3 *pol2-M644G* plasmid), pYTK060 (type 4a *3xFLAG-6xHis* plasmid), pYTK066 (type 4b *tTDH1* plasmid), 1x T4 DNA ligase reaction buffer (NEB, B0202), 1,500 units of T7 DNA ligase (NEB, M0318), 10 units of *Bsa*I-HFv2 (NEB, R3733), Milli-Q H_2O to a final volume of 10 µL. The assembly reaction and bacterial transformation were performed as above. The pYTK096 vector conferred kanamycin

resistance to successful transformants, so following recovery in LB medium, cells were plated onto LB medium supplemented with kanamycin.

2.6.3 Sanger sequencing

Sanger sequencing performed by Source BioScience was used to confirm the sequences of plasmids and PCR amplicons of genotypes of interest.

2.6.4 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate and analyse DNA molecules based on their size. Agarose powder (SeaKem LE agarose, Lonza, 50004) was combined with either a neutral buffer for native gel electrophoresis or an alkaline buffer for denaturing gel electrophoresis (see below). Electrophoresis was carried out in Bio-Rad (Wide) (Mini) Sub Cell GT tanks, depending on the size of the gel, with the voltage provided by a Bio-Rad PowerPac Basic or HC. To enable estimation of sample sizes, markers of known size were run on each gel (typically Quick-Load 2-Log DNA ladder (NEB, N0469S) or HyperLadder 1 kb (Bioline, BIO-33053) were used). Gels were generally post-stained with ethidium bromide (Sigma, E1510) diluted to a final concentration of 3 µg/mL in the appropriate buffer or 1x SYBR Gold Nucleic Acid Stain (Invitrogen, S11494) in the appropriate buffer. Ethidium bromide-stained gels were rinsed in H_2O prior to visualisation. Standard gel images were captured using a Benchtop UV Transilluminator (UVP BioDoc-It Imaging System). For the detection of labels or stains at different wavelengths to those of ethidium bromide and SYBR Gold, a PharosFX Molecular Imager System with External Laser (Bio-Rad) was used.

<u>Native gel electrophoresis</u>

For most native gels, agarose was suspended in TBE buffer diluted to 1x in distilled H_2O from a 10x stock solution (892 mM Trizma base (Sigma, T6066), 890 mM boric acid (Fisher Chemical B/3800/53), 20 mM EDTA, pH 8.0 (VWR, E177)) and heated until dissolved. Purple Gel Loading Dye (NEB, B7024) was added to a final concentration of 1x to DNA samples and

mixed prior to loading. Gels were run in 1x TBE buffer and post-staining was carried out in DNA stain diluted in 1x TBE.

In some instances, for example for analysis of larger DNA fragments, TAE buffer was used to make, run and stain gels. A 1x working concentration was obtained by diluting a 50x stock (2 M Trizma base (Sigma, T6066), 1 M acetic acid (Sigma, A6283), 50 mM EDTA, pH 8.0 (VWR, E177)) in distilled H_2O.

Denaturing gel electrophoresis

Denaturing gels were made by dissolving agarose powder in distilled H_2O by heating. Once the molten agarose had cooled to approximately 60 °C, alkaline electrophoresis buffer was added to a final concentration of 1x from a 10x stock. Alkaline electrophoresis buffer (10x) had the following composition: 10 mM EDTA, pH 8.0, 500 mM NaOH in distilled H_2O and was freshly made for each run.

For DNA backbone nicking at ribonucleotides in polymerase-usage strains and strand pre-denaturation, 5 µg of DNA per lane were supplemented with 0.3 M NaOH and heated at 55 °C for 2 h prior to denaturing gel electrophoresis.

Samples and DNA ladders were combined with alkaline loading buffer prior to loading. Alkaline loading buffer (6x) was prepared for each run by mixing equal volumes of the two solutions:
Solution 1 = 12 mM EDTA, 600 mM NaOH, 30% (w/v) Ficoll 400
Solution 2 = 0.5% xylene cyanol FF, 0.3% Orange G
and added to a final concentration of 1x to DNA samples.

Gels were run in alkaline running buffer (1 mM EDTA, pH 8.0, 50 mM NaOH) in Bio-Rad Sub Cell GT tanks at 30 V for at least 26 h. It was necessary to neutralise gels prior to DNA staining using ethidium bromide. After electrophoresis, gels were rinsed for a few minutes in H_2O, then

twice in neutralising buffer (1 M Trizma base, 1.5 M NaCl, pH adjusted to 7.4 using 37% HCl) each for 15 min. Gels were subsequently stained with ethidium bromide diluted in TAE.

Pulsed-field gel electrophoresis (PFGE)

PFGE was used for the size separation of ultra-long molecules (in the order of hundreds of kb). A Bio-Rad Chef Mapper XA, Bio-Rad Mini Chiller, and Bio-Rad Variable Speed Pump were used for PFGE, according to the manufacturer's instructions. CHEF DNA Size Marker, 0.2-2.2 Mb, *S. cerevisiae* Ladder (Bio-Rad, 1703605) plugs were run alongside samples to allow estimation of sample sizes. Optimal conditions were investigated as described in the relevant Results chapter sections. In general, 1% (w/v) agarose powder was dissolved in 150 mL of 0.5x TBE buffer for each gel. DNA samples embedded in agarose plugs were sealed in wells with agarose. Unused wells were also sealed with agarose. Prior to a run, the gel chamber was filled with approximately 2.1 L of 0.5x TBE buffer which was circulated through the cooling module to allow pre-chilling to 14 °C. The variable speed pump was set to approximately 70 (corresponding to a circulation rate of ~1 L/min). The buffer circulation was stopped to allow the placement of the gel within the chamber. Samples in solution, rather than agarose plugs, were carefully loaded using a wide-bore pipette tip at this stage. If such samples were loaded, the gel was run for 15-20 min prior to restarting the buffer circulation to allow the DNA to enter the gel. The run times were varied as described in the relevant Results sections to find the optimal run duration. The electrophoresis conditions were as follows: voltage = 6 V/cm, angle = 120 °, pulse time = 60-120 s, ramped linearly during the run, buffer temperature = 14 °C, buffer circulation by speed pump set to 70. Gels were post-stained with ethidium bromide diluted in 0.5x TBE.

2.6.5 Southern blotting

Southern blotting experiments began with either denaturing or native TAE gel electrophoresis (section 2.6.4). After DNA staining and visualisation, gels were subjected to acid nicking in 0.25 M HCl for 10 min. They were then rinsed in distilled H_2O on a 2D horizontal rocker for 10 min, followed by denaturation in 1.5 M NaCl, 0.5 M NaOH first for 5 min, then for 40 min in

fresh buffer. The gel was then treated with neutralising buffer (1 M Trizma base, 1.5 M NaCl, pH to 7.4 with 37% HCl) for two 15 min incubations. DNA transfer was performed under neutral conditions under vacuum. For each transfer, a 15 x 25 cm rectangle of neutral nylon membrane (MP Biomedicals, 11MEMN0001) was pre-wetted in distilled H_2O for 5 min before being placed on the surface of some 20x SSC (3 M NaCl, 0.3 M trisodium citrate, sterilised by autoclaving). Vacuum transfers were assembled and performed using VacuGene XL Vacuum Blotting System (GE Healthcase Bio-Sciences) according to manufacturer's instructions. In brief, the porous support screen was pre-wetted in distilled H_2O and placed in the blotting unit. The equilibrated transfer membrane was placed on the support screen. Polyethylene masks were cut to have an ~0.5 cm overlap with the gel around the edge. The mask was placed on top of the membrane and any air bubbles between the membrane and mask were carefully removed. The top frame was placed on the blotting unit but not yet clipped into place. The neutralised gel was carefully lifted from the buffer and placed onto the membrane, ensuring overlap with the mask on all edges and avoiding air bubbles between the membrane and the gel. The vacuum pump was switched on before the top frame was clamped in place. Approximately 500 mL of 20x SSC were slowly poured onto the gel to fill the blotting unit, with the gel fully submerged. The vacuum was set to approximately 50 mbar and the transfer was left to run for 1 h. At the end of the transfer, the blotting unit was tilted using the flip-up stand and excess 20x SSC was removed using the vacuum tube. The top frame, gel, and mask were removed before the vacuum pump was switched off. The membrane was carefully removed, placed on Whatman paper and transferred DNA was crosslinked to the membrane by 120 mJ/cm^2 ultraviolet (UV) light treatment (Ultraviolet Crosslinker, Cole-Parmer). The gel was rinsed in H_2O, re-stained and visualised to monitor the transfer efficiency.

To synthesise probes complementary to the region of interest, 100 µL PCR reactions were used to amplify the target sequence. PCR products were run on a native gel and extracted from the gel using a QIAquick Gel Extraction Kit (Qiagen, 28704) according to the manufacturer's instructions. For the *rnh202-HBD* genotyping Southern (Figure 4.9), the purified PCR product was combined with 1 µL each of 2-Log ladder and HyperLadder 1 kb that

had been purified from their buffers using a GenElute PCR Clean-Up kit according to the manufacturer's instructions. The volume was adjusted to 16 µL by SpeedVac and the DNA was denatured by incubation at 95 °C for 10 min. Then, 4 µL of Fluorescein-High Prime labelling mixture (Roche, 11 585 622 910) were added to the denatured DNA and the mix was incubated at 37 °C for 20 h in a Bio-Rad Tetrad thermal cycler. Following the 20 h incubation, the reaction was stopped by heating to 65 °C for 10 min.

To synthesise strand-specific probes for the assessment of ribonucleotide incorporation in polymerase-usage strains (Figure 4.7), the concentration of the gel-purified PCR template was assessed spectrophotometrically. The template was diluted to 1 ng/µL with Milli-Q H_2O and the following 25 µL reactions were set up for each probe: 1 ng of template DNA, 1x NH_4 reaction buffer (Bioline), 3 mM magnesium chloride (Bioline), 100 µM each of dATP, dCTP, and dGTP, 39 µM dTTP, 21 µM Fluorescein-12-2'-dUTP (Roche, 11 373 242 910), 0.5 units of BIOTAQ DNA polymerase (Bioline), 10 pmol of oligonucleotide primer (EH216 or EH217), Milli-Q H_2O to final volume of 25 µL. Reactions were incubated in a Bio-Rad Tetrad thermal cycler with the following programme: initial denaturation at 94 °C for 5 min, then 25 cycles of 94 °C for 1 min, 55 °C for 30 s, 72 °C for 1 min, followed by a final extension step at 72 °C for 5 min. To synthesise a HyperLadder 1 kb probe, the HyperLadder 1 kb DNA was purified using a GenElute PCR Clean-Up kit (Sigma, NA1020) according to manufacturer's instructions. The resulting DNA was diluted 0.5 µL into 7.5 µL of Milli-Q H_2O, denatured at 95 °C for 10 min, then incubated with 2 µL of Fluorescein-High Prime kit as above.

After crosslinking, the membrane was incubated in a hybridisation tube with 50 mL of pre-warmed hybridisation buffer (5x SSC, 0.1% (w/v) SDS, 5% (w/v) dextran sulphate (Sigma, D8906), 0.2% (w/v) I-Block protein-based blocking reagent (Applied Biosystems, T2015)) at 60 °C for 2 h with tube rotation in a Hybrigene HB-3D hybridisation incubator (Cole Palmer). DNA probes were denatured at 95 °C for 5 min, snap-cooled, then added to the hybridisation buffer for incubation at 60 °C overnight. The next day, the membrane was rinsed three times in 100 mL of wash buffer 1 (0.1% (w/v) SDS, 1x SSC) at 60 °C each for 10 min, then three

times in 100 mL of wash buffer 2 (0.1% (w/v) SDS, 0.5x SSC) at 60 °C each for 10 min. It was then transferred to a tray for a brief rinse in 1x AB buffer (100 mM Trizma base, 150 mM NaCl, pH adjusted to 7.55 with 37% HCl). This and all subsequent steps were performed at room temperature. After the AB rinse, the membrane was incubated in 250 mL of 1% blocking buffer (1% (w/v) skimmed milk powder in 1x AB buffer) on a 2D horizontal rocker for 1 h. The blocking buffer was then replaced with 250 mL of 0.5x blocking buffer containing 0.75 units of anti-fluorescein-AP Fab fragment (Roche, 11426338910) and the membrane was incubated for 1 h. The membrane was then rinsed three times in 500 mL of 1x AB buffer containing 0.2% (v/v) Tween-20, each for 10 min. After rinsing, excess liquid was drained from the membrane as it was transferred to a plastic sheet. For visualisation, 2 mL of CDP-Star Detection Reagent (GE Healthcare Life Sciences, RPN3682) were applied to the membrane and left to react with the alkaline phosphatase conjugated to the anti-fluorescein antibody fragments for 30 min in the dark. After this time, the membrane was transferred to a fresh plastic sheet and the signal detected using an ImageQuant LAS 4000 imager set to detect chemiluminescence, with the membrane on tray 4 and incremental exposure with high sensitivity at 15 min interval times.

2.6.6 Densitometry analysis of gel and Southern membrane images

To obtain plots of signal intensity in gel or Southern lanes, image TIFF files were opened in ImageJ and a box was drawn around a lane of interest. The contents of the rectangle were analysed by going to *Analyse>Gels>Plot Lane* and the resulting curve was selected with the *Wand* tool. This allowed removal of the background by selecting *Edit>Clear Outside*. Next, *Analyse>Tools>Analyse Line Graph* was used, at which point the x and y coordinates of the plot could be saved as a csv file. The rectangle was shifted to the next lane without adjusting its size or vertical position on the image and the process was repeated for all lanes in an image. The csv files were then opened in `R studio` and the data from all lanes were combined into a single data frame to facilitate plotting. Plots were produced using the `ggplot2` package and the data were smoothed using the Loess method with the formula 'y ~ x' called by the `stat_smooth()` command where indicated. The data for each lane of the Southern blot image were additionally normalised to the area under the curve prior to plotting.

2.6.7 Fluorometric measurement of DNA concentration

For general assessment of the concentration of DNA samples, the Qubit dsDNA HS Assay Kit (Invitrogen, Q32854) was used as follows. A master mix of 200 µL Qubit dsDNA HS buffer with 1 µL of Qubit dsDNA HS dye per sample and standard was made. Depending on the predicted approximate amount of DNA, 1-20 µL of DNA sample were added to the Qubit dsDNA HS buffer and dye master mix to a final volume of 200 µL in Qubit Assay tubes. For the DNA standards, 10 µL of each standard were added to 190 µL of the buffer and dye master mix. Samples were mixed by vortexing and incubated at room temperature in light-protected conditions to prevent photobleaching of the dye for at least 5 min before measurement using a Qubit fluorometer or DeNovix fluorometer. For assessment of DNA concentration and homogeneity following plug lysis purification, the Qubit dsDNA BR Assay Kit (Invitrogen, Q32853) was used to measure 2 µL aliquots from the top, middle, and bottom of each tube.

2.6.8 Spectrophotometric measurement of DNA concentration and purity

The concentration and purity of DNA samples were measured spectrophotometrically using either a DeNovix DS-11 spectrophotometer, a NanoDrop Lite spectrophotometer, or a NanoDrop ND-1000 spectrophotometer. The A260/A230 and A260/A280 ratios were used to assess contamination from residual chemicals or protein, respectively.

2.6.9 Ethanol precipitation of DNA

To precipitate DNA, 3 M sodium acetate (pH 5.2) was added to a final concentration of 0.3 M, followed by 2-3 volumes of absolute ethanol. Samples were mixed by inversion and centrifuged at 15,871 x *g* for 15 min. The DNA pellet was rinsed in 70% (v/v) ethanol and allowed to air dry. The DNA was resuspended in TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) or Milli-Q H_2O, depending on subsequent steps.

2.6.10 Solid Phase Reversible Immobilisation (SPRI) bead purification of DNA

For SPRI bead purification and size selection of DNA, AMPure XP (Beckman Coulter) beads were used. Beads were equilibrated to room temperature for 30 min and vortexed for 30 s

before each use. The appropriate volume of beads, depending on sample volume and desired bead:sample ratio, were added and the sample was vortexed and incubated on a tube rotator for 10-15 min at room temperature. Tubes were pulse-centrifuged and placed in a magnetic rack to collect the beads at one side of the tube. The supernatant was removed and the beads were rinsed twice with 200 µL of 80% (v/v) ethanol. After the second wash, tubes were pulse-centrifuged and placed back in the magnetic rack and residual ethanol was removed. Tubes were left open to allow the beads to air-dry. Once the ethanol had been removed, DNA was eluted from the beads by addition of TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) or Milli-Q H_2O, vortexing and incubation at 37 °C in a Thermomixer comfort (Eppendorf) at 500 rpm for at least 30 min. The tubes were pulse-centrifuged and placed in a magnetic rack so that the DNA-containing eluate could be transferred to a fresh 1.5 mL tube.

2.6.11 dUTP-Cy3 incorporation trials

For the aminoallyl-dUTP-Cy3 (referred to as dUTP-Cy3) incorporation during PCR trial (Figure 5.5), PCRs were set up with varying proportions of dUTP-Cy3 and dTTP as indicated in the figure. The reactions were performed in 1x ThemoPol Reaction Buffer (NEB, B9004), with 200 µM each of dATP, dCTP, and dGTP, and dUTP-Cy3 and dTTP added in varying proportions to a final combined concentration of 200 µM, 500 nM of each primer, 0.1 ng of template plasmid DNA, and approximately 0.125 units of *Taq* DNA polymerase (NEB, M0267) in each 10 µL reaction. PCR was performed in a Bio-Rad Tetrad thermal cycler with the following programme: initial denaturation at 95 °C for 30 s, a total of 35 cycles of 95 °C for 30 s, 54 °C for 30 s, 68 °C for 30 s, followed by a final incubation at 68 °C for 5 min. the PCR products were purified using a GenElute PCR Clean-Up kit and analysed by native agarose gel electrophoresis.

The nick-label trial in Figure 5.6 was performed on an amplicon produced by standard PCR and purified using a GenElute PCR Clean-Up kit. A starting amount of 1.5 µg of DNA per lane was used. The DNA was nicked in a 60 µL reaction comprising: 9 µg of DNA (for six lanes), 1x NEBuffer 3.1 (NEB, B7203), 10 units of Nt.*Bsp*QI (NEB, R0644), and Milli-Q H_2O. The

reaction was incubated at 50 °C for 1 h before dilution with 60 µL of Milli-Q H_2O. Labelling reactions were performed in a final volume of 40 µL each, with 25 µM of each dNTP (dUTP-Cy3 and dTTP in varying proportions to a final combined concentration of 25 µM), 1x ThermoPol Reaction Buffer, 0.2 units of *Taq* DNA polymerase (NEB). Reactions were incubated at 72 °C for 1 h, then the DNA was ethanol precipitated.

For dUTP-Cy3 labelling of 2-Log DNA ladder, 5 µL of ladder per lane were diluted in Milli-Q H_2O for purification using a GenElute PCR Clean-Up kit. End-labelling reactions were set up as follows for one lane's worth of labelled ladder: 1x NEBuffer 2 (NEB, B7002), purified ladder DNA, 33 µM dUTP-Cy3, 33 µM each of dATP, dCTP, and dGTP, 1 unit of DNA polymerase I, Large (Klenow) Fragment (NEB, M0210) and Milli-Q H_2O to a final volume of 30 µL. Reactions were incubated at 25 °C for 15 min, before addition of 10 mM EDTA and incubation at 75 °C for 20 min to stop the reaction. The end-labelled DNA ladder was ethanol precipitated before use in gel electrophoresis.

2.6.12 Fluorescent labelling of OFs by nick translation and OF repair

This was a novel approach, and so a standard protocol was not available. Therefore, the precise protocol changed from one experiment to another in an attempt to optimise the conditions, as described in the relevant sections of the Results chapter.

For estimation of labelled genomic DNA smears run on a denaturing gel, 2-Log DNA ladder was end-labelled as described in section 2.6.11.

Genomic DNA was labelled with a starting amount of 4 µg of DNA per gel lane. The initial trial (Figure 5.8) was performed using 35% dUTP-Cy3 : 65% dTTP, with a combined final concentration of 25 µM, 25 µM each of dATP, dCTP, and dGTP, 1x ThermoPol reaction buffer, 2 units of *Taq* DNA polymerase (NEB) in a total reaction volume of 40 µL. Tubes were incubated in a Bio-Rad Tetrad thermal cycler at 72 °C for 1 h, then ethanol precipitated prior to running on a 1.5% denaturing agarose gel.

The next trial (Figure 5.10) was performed with 100% dUTP-Cy3 (no dTTP) but lower dNTP concentrations, based on labelling reactions in the literature (McCaffrey et al. 2017; De Carli et al. 2016; Gabrieli et al. 2017, 2018b; Sharim et al. 2019). Labelling reactions were performed in a final volume of 100 µL, each containing 4 µg of DNA, with 1x ThermoPol Reaction Buffer, 20 units of *Taq* DNA polymerase, and 2.5 µM of each dNTP (with dTTP replaced by dUTP-Cy3). Reactions were incubated as above.

For the repair of labelled OFs, 125 µL reactions were set up using the 100 µL of labelled DNA from the previous step with 300 units of *Taq* DNA ligase (NEB, M0208), 100 µM of each dNTP, 1 mM NAD^+, 2.5 µL additional 10x ThermoPol Reaction Buffer (to account for the extra 25 µL). Reactions were incubated in a Bio-Rad Tetrad thermal cycler at 45 °C with the lid set to 55 °C for 30 min. Samples were then ethanol precipitated prior to denaturing agarose gel electrophoresis.

For the OF labelling and repair for the Saphyr experiment, 1 µg of genomic DNA was used as the input for each reaction. The reagent and enzyme amounts were adjusted accordingly. In brief, 5 units of *Taq* DNA polymerase (NEB) and 625 nM of each dNTP were used. Rather than dUTP-Cy3, aminoallyl-dUTP-ATTO647N was used for OF labelling. Reaction volumes were 23 µL and were incubated as above, with the lid of the thermal cycler set to 82 °C (rather than “ON”). The repair reactions were performed using 140 units of *Taq* DNA ligase (NEB) for 1 µg of DNA in 30 µL reaction volumes as above. The labelled and repaired DNA was then prepared for Saphyr analysis (section 2.7.1).

2.6.13 Measurement of DNA molecule lengths by the Agilent 4200 TapeStation System

To ensure optimal clustering during a NextSeq sequencing run, the DNA fragments of approximately 300-500 bp in length were required. This was checked using the Agilent TapeStation system; an electrophoresis-based method of determining DNA fragment size. Before the analysis, the High Sensitivity D1000 reagents were allowed to equilibrate to room temperature for 30 min. Reagents were vortexed and pulse-centrifuged before 2 µL of High

Sensitivity D1000 Sample Buffer per sample and ladder were pipetted into PCR strips. Next, 2 µL of ladder or DNA sample were pipetted into the sample buffer, samples were mixed and pulse-centrifuged. The High Sensitivity D1000 ScreenTape device was checked for bubbles and loaded into the TapeStation. Samples were loaded into the ScreenTape by the TapeStation and analysed. Average fragment size was estimated from regions drawn around the main distribution of fragment sizes in the Agilent Tapestation Analysis Software.

2.6.14 Measurement of Illumina NextSeq library concentration by quantitative PCR (qPCR)

The concentrations of Illumina NextSeq libraries had to be accurately quantified to ensure appropriate library pooling for sequencing. This was achieved using qPCR-based quantification of library dilutions. NEBNext Library Quant Primer Mix was combined with NEBNext Library Quant Master Mix at a ratio of 1:15. NEBNext Library Quant Dilution Buffer was diluted to 1x in Milli-Q H_2O, allowing 1.2 mL per library to be measured. Libraries were serially diluted in 1x Dilution Buffer to 1:10,000 – 1:250,000 (depending on the estimated library concentration from TapeStation analysis). For each sample and standard, 16 µL of the Library Quant Master Mix with Primers and 4 µL of DNA were pipetted into 0.1 mL 4-strip Rotor-Gene Style tubes (STARLAB) in triplicate. No template controls had 4 µL of 1x Dilution Buffer added. Samples were analysed in a Qiagen Rotor-Gene Q real-time PCR cycler using the SYBR setting with the following cycling conditions: 95 °C for 1 min, 95 °C for 15 s, 63 °C for 45 s, cycling back to the second step 34 times, with fluorescence recorded after each 63 °C step. The concentration of each sample was calculated based upon the Ct value and average fragment size (from the TapeStation measurement). Multiple libraries were pooled to a final concentration of 4 nM.

2.7 Genomic methods

2.7.1 Preparation of genomic DNA for Irys or Saphyr analysis

Samples were prepared for Irys analysis by performing Nick Label Repair Stain (NLRS) reactions and for Saphyr analysis by Direct Label Stain (DLS) reactions according to protocols

from BNG. In both instances, the starting material was genomic DNA purified by plug lysis (section 2.4.8) that had equilibrated to room temperature for 30 min.

NLRS labelling required a starting DNA concentration of 35 – 200 ng/µL. A total of 300 ng of DNA was used in each NLRS sample. This was first treated with the nicking endonuclease Nt.*Bsp*QI, which had been determined to have the appropriate frequency of recognition sites in both the *S. cerevisiae* and *S. pombe* reference genomes using BNG's LabelDensityCalculator and Knickers software. The following reactions were prepared in 200 µL thin-wall PCR tubes on ice:

Component	**Amount**	**Reagent source**
Genomic DNA	300 ng	Experimental sample
10x Buffer 3	1 µL	BNG (NLRS kit part 20251)
10 U/µL Nt.*Bsp*QI	1 µL	NEB (R0644S)
Ultrapure H_2O	To final volume of 10 µL	BNG (NLRS kit part 20165)

The reactions were gently mixed by pipetting, then incubated in a thermal cycler at 37 °C, with the lid at 47 °C, for 2 h. Tubes were briefly pulse centrifuged to remove any condensation from the tube walls and placed on ice for the labelling reactions:

Component	**Amount**	**Reagent source**
Nicked genomic DNA	300 ng in 10 µL	Nicking reactions
10x Labelling Buffer	1.5 µL	BNG (NLRS kit part 20168)
10x Labelling Mix	1.5 µL	BNG (NLRS kit part 20167)
5 U/µL *Taq* DNA polymerase	1 µL	NEB (M0267S)
Ultrapure H_2O	1 µL	BNG (NLRS kit part 20165)

Reactions were gently mixed and briefly pulse centrifuged before incubation in a thermal cycler at 72 °C, with the lid at 82 °C, for 1 h. Tubes were briefly pulse centrifuged and placed on ice for the repair reactions:

Component	Amount	Reagent source
Nick-labelled genomic DNA	300 ng in 15 µL	Nick and label reactions
10x ThermoPol Reaction Buffer	0.5 µL	NEB (provided with M0267S)
50x Repair Mix	0.4 µL	BNG (NLRS kit part 20250)
50 mM NAD^+	0.4 µL	NEB (B9007S)
40 U/µL *Taq* DNA ligase	1 µL	NEB (M0208S)
Ultrapure H_2O	2.7 µL	BNG (NLRS kit part 20165)

Reactions were gently mixed and briefly pulse centrifuged before incubation in a thermal cycler at 37 °C, with the lid at 47 °C, for 30 min. Tubes were briefly pulse centrifuged. The staining mix was prepared at room temperature in a 0.6 mL amber tube with minimal exposure to light:

Component	Amount	Reagent source
Nick-labelled-repaired genomic DNA	300 ng in 20 µL	Nick, label, repair reactions
Stop Solution	1 µL	BNG (NLRS kit part 20164)
4x Flow Buffer	15 µL	BNG (NLRS kit part 20161)
5x DTT	12 µL	BNG (NLRS kit part 20166)
DNA Stain	3 µL	BNG (NLRS kit part 20162)
Ultrapure H_2O	9 µL	BNG (NLRS kit part 20165)

Reactions were gently mixed by pipetting up and down with a wide-bore pipette tip, briefly pulse centrifuged, and incubated at 4 °C overnight to achieve uniform staining. The NLRS DNA was fluorometrically quantitated with the Qubit dsDNA High Sensitivity Assay Kit (section 2.6.7) to ensure it fell within the target concentration range of 3-10 ng/µL prior to loading.

Irys chips were loaded according to instructions from BNG. For the *S. cerevisiae* run, training and supervision was provided by John Baeten of BNG for NLRS, flow cell loading and starting a run. In brief, a chip was allowed to equilibrate to room temperature for 30 min before loading. The chip was loaded into the Irys machine and 8 µL of the NLRS DNA were carefully pipetted into the Inlet well of a flow cell, followed by 8 µL in the Outlet well. Oil was applied to the objective, then the machine detected the chip and correct objective focussing was verified. A "recipe" for loading of each flow cell was determined during user guided loading steps to ensure optimal sample loading in each flow cell. The recipes for the two runs were:

	***S. cerevisiae* run**	***S. pombe* run**
Concentration time (s)	400	310
Concentration voltage (V)	30	30
Optimised bump time (s)	14	20
Optimised load time (s)	32	36

Data were collected during 26 (*S. cerevisiae*) or 30 (*S. pombe*) scans, and run reports provided information about raw molecules that had been pre-filtered to include only molecules ≥ 20 kb in length, and with a label signal-to-noise ratio (SNR) ≥ 3. The SNR is defined as the ratio of brightness of the nick labels when compared to background noise.

DLS labelling required a starting DNA concentration of 36-150 ng/µL. Each DLS reaction had the following composition:

Component	**Amount**	**Reagent source**
Genomic DNA	750 ng	Experimental sample
5x DLE-1 Buffer	6 µL	BNG (DLS kit part 20350)
20x DL-Green	1.5 µL	BNG (DLS kit part 20352)
20x DLE-1 Enzyme	1.5 µL	BNG (DLS kit part 20351)
Ultrapure H_2O	To final volume of 30 µL	BNG (DLS kit part 20355)

Reactions were set up with minimal light exposure in 200 µL thin-wall PCR tubes, with addition of the genomic DNA and ultrapure H_2O first. When multiple samples were being labelled, a labelling master mix was made, 9 µL of which were added to each sample. Reactions were gently mixed and briefly pulse centrifuged before incubation in a thermal cycler at 37 °C, with the lid at 47 °C, for 2 h. Immediately following this incubation, the tubes were pulse centrifuged, then 5 µL of Puregene Proteinase K (QIAGEN, 158920) were added to each sample. The tubes were incubated in a thermal cycler at 50 °C, with the lid at 60 °C, for 30 min. Subsequently, samples underwent two sequential drop dialyses as follows. First, one DLS membrane (13 mm, DLS kit part 20358) per sample was pre-wetted with 25 µL of 1x DLE-1 buffer (5x DLE-1 buffer (DLS kit part 20350) diluted in ultrapure H_2O (DLS kit part 20355)) in a well of a 24 well plate (DLS kit part 20357). Immediately after the proteinase K treatment, the labelled DNA sample was applied to the centre of the pre-wetted membrane and the well was

sealed using a plate sealing strip (DLS kit part 20361) to prevent evaporation. The plate was covered by a cardboard box to protect samples from light and the samples were incubated at room temperature for 1 h. After the first drop dialysis, the sample was then transferred to a second pre-wetted membrane and incubated for 30 min. For the staining, 20 µL of the labelled sample were transferred from the membrane to a 2 mL round bottom amber tube (DLS kit part 20362). Any residual sample was transferred to a separate tube and stored at 4 °C. A staining master mix was made using the following recipe (volumes given are for one sample):

Component	**Volume (µL)**	**Reagent source**
4x Flow Buffer	15	BNG (DLS kit part 20353)
1 M DTT	6	BNG (DLS kit part 20355)
DNA stain	3.5	BNG (DLS kit part 20356)
Ultrapure H_2O	15.5	BNG (DLS kit part 20355)

The master mix was vortexed and pulse centrifuged before 40 µL were added to the top of each DNA sample. For further light protection, amber tubes were wrapped in aluminium foil during incubation at room temperature on a tube rotator set to 5 rpm for 1 h to mix the staining solution with the DNA. Following the 1 h mixing, the 2 mL amber tubes were pulse centrifuged and samples were incubated at room temperature protected from light overnight.

The Saphyr chip was loaded according to the manufacturer's instructions with guidance and supervision from Dr. Barbara Xella.

2.7.2 5-Bromo-2'-deoxyuridine (BrdU) immunoprecipitation (IP) and preparation of libraries for Illumina NextSeq 550 sequencing

BrdU-containing DNA was immunoprecipitated and sequenced using a modified protocol adapted from (Viggiani, Knott and Aparicio, 2010). Following extraction and purification of genomic DNA (section 2.4.8), the concentration of each DNA sample was measured (section 2.6.7). DNA was then fragmented by sonication using a Diagenode Bioruptor UCD-200 at 4 °C. For fragmentation, each DNA sample was made to a total volume of 400 µL with 1x TE in 1.5 mL tubes and the tube holder was filled with tubes to ensure homogeneous sonication.

Sonication was performed at low intensity at 30 s on/off intervals for a total of 20 cycles to shear DNA to 300-500 bp. Fragmented DNA was precipitated by ethanol precipitation (section 2.6.9) and resuspended in 25 µL of TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). The DNA concentration was again measured using the Qubit dsDNA HS assay kit. When multiple samples were being processed for subsequent pooling, the amount of each sample carried forward was adjusted to approximately that of the lowest amount available. To enable pooling of samples before BrdU-IP, sample-specific adaptors needed to be ligated so that different samples could be distinguished during data analysis. First, the ends of DNA molecules needed to be modified to allow adaptor ligation by addition of a dA-tail. This reaction required 25 µL of the fragmented DNA (the total amount after sonication and precipitation), 3.5 µL of NEBNext Ultra II End Prep Reaction Buffer, and 1.5 µL of NEBNext Ultra II End Prep Enzyme Mix. The reactions were incubated in a thermal cycler at 20 °C for 30 min, followed by 65 °C for 30 min (lid set to 100 °C). The adaptor ligation reactions were set up immediately following the End Prep reactions to minimise loss of the added dA-tails. Adaptor and index oligonucleotide sequences were taken from (Dunham and Friesen 2013). Multiplexing Adaptors were formed by mixing equal volumes of 100 µM oligonucleotides from each adaptor pair with 10x NEBuffer 2 diluted to a final concentration of 1x (final concentration of each oligonucleotide was 45 µM). The oligonucleotides were annealed using the following programme in a thermal cycler: 95 °C for 5 min, 90 °C for 1 min, 85 °C for 1 min, 80 °C for 1 min, 75 °C for 30 s, then dropping 1 °C every 30 s for 69 cycles and holding at 4 °C or on ice for 30 min before use. Adaptor ligation reactions consisted of the following: all of the dA-tailed DNA (30 µL), 15 µL of NEBNext Ultra II Ligation Master Mix, 0.5 µL of NEBNext Ligation Enhancer, 1.25 µL of 45 µM Multiplexing Adaptor. Reactions were incubated at 20 °C for 25 min in a thermal cycler with the lid “OFF”. Adaptor-ligated DNA was purified using SPRI beads (section 2.6.10) added to 0.9x and eluted in 20 µL of Milli-Q H_2O. Then the concentration of each sample was measured using the Qubit dsDNA HS assay kit so that the same amount of each sample could be pooled for the IP step. A total of 1-3 µg of pooled DNA were used in each BrdU-IP. After pooling the samples, ~20 ng of the input DNA were transferred to 20 µL Milli-Q H_2O in a 1.5 mL tube and stored at 4 °C until the indexing and amplification step. The remaining DNA was vacuum concentrated to a

volume of 45 µL using a Genevac miVAC DNA Vacuum-Integrated Centrifugal Concentrator System. PBS was added to a final concentration of 1x. The DNA was denatured by heating at 95 °C for 10 min and snap-cooling in icy water. The sample was collected at the bottom of the tube by pulse-centrifugation. For the BrdU-IP, 185 µL of ice-cold IP buffer (1x PBS, 0.0625% (v/v) Triton X-100) were added, followed by 1.5 µg/mL of anti-BrdU antibody (15 µL of 25 µg/mL Purified Mouse Anti-BrdU, Clone B44, BD Biosciences). Samples were incubated overnight at 4 °C on a tube rotator. Then, samples were pulse-centrifuged before addition of 30 µL of a 50% suspension of Dynabeads Protein G (Invitrogen) in PBS using a wide-bore pipette tip. Tubes were incubated at 4 °C for 1 h on a tube rotator, pulse-centrifuged and placed in a magnetic rack. The supernatant was removed and the beads were washed three times with 1 mL of ice-cold IP buffer by adding the buffer to the tube and incubating on a tube rotator at room temperature for 3 min. The IP buffer was removed after pelleting the beads in the magnetic rack. After the third IP buffer wash, the beads were washed in 1 mL of TE (pH 7.6). DNA was eluted from the beads by addition of 100 µL of elution buffer (TE (pH 7.6) containing 1% (w/v) SDS) and incubation at 65 °C in a Thermomixer comfort (Eppendorf) at 500 rpm for 15 min. The eluate was transferred to a fresh 1.5 mL tube after pelleting the beads on a magnetic rack. The DNA was purified using 0.9x AMPure XP beads and eluted in 24 µL of H_2O. The input and IP DNA samples were then amplified using indexing primers by combining: 23.5 µL of DNA, 25 µL of NEBNext Ultra II Q5 Master Mix, 0.25 µL of 100 µM Multi/Std PCR Primer 1.0, 0.25 µL of 100 µM sample-specific index primer, and 1 µL of 2.5 µM multiplexing PCR primer 2.0 for each reaction. Reactions were incubated in a thermal cycler with the lid set to 100 °C using the following programme: 98 °C for 30 s, 98 °C for 10 s, 65 °C for 75 s (steps 2-3 for 18 cycles), 65 °C for 5 min, hold at 4 °C. The indexed DNA was then purified by two sequential AMPure bead purifications using 0.9x beads. For the final elution step, the DNA was eluted in 20 µL of H_2O.

2.7.3 Illumina NextSeq 550 DNA sequencing

Sequencing runs were set up with guidance and supervision from Amanda Williams. Sodium hydroxide stock solution was diluted to 200 mM using buffer PW1 (pure water). To denature

the library, 5 µL of 200 mM NaOH were added to 5 µL of the 4 nM library, followed by 5 min incubation at room temperature. To neutralise the NaOH, 5 µL of 200 mM Tris-HCl (pH 7.0) were added, followed by 985 µL of Hybridisation Buffer (HT1). The resulting 20 pM library was diluted to 2.2 pM using HT1 in a final volume of 1.2 mL. The library was loaded in the reagent cartridge and analysed using a NextSeq 550 for 80 cycles.

2.7.4 Preparation of ONT sequencing libraries and flow cell loading

Genomic DNA purified from *S. pombe* (section 2.4.8) was prepared for MinION sequencing using the 1D Genomic DNA by Ligation (SQK-LSK109) kit according to the manufacturer's instructions, with some minor modifications. Where possible, approximately 5 µg of DNA was used for the preparation (or as much as was available if < 5 µg). The DNA was transferred to a DNA LoBind 1.5 mL tube (Eppendorf, 0030108051) and the volume was made up to 48 µL with Milli-Q H_2O. The following reagents were added to each DNA sample:

Component	**Volume (µL)**	**NEB product number**
NEBNext FFPE DNA repair buffer	3.5	E6622
NEBNext FFPE DNA repair mix	2.0	M6630
NEBNext Ultra II End-prep reaction buffer	3.5	E7647
NEBNext Ultra II End-prep enzyme mix	3.0	E7646

Tubes were gently flicked to mix, then incubated at 20 °C for 30 min, followed by 65 °C for 30 min. This repaired common forms of DNA damage, such as nicks and gaps, oxidised bases, and cytosine deamination to uracil. In addition, 5' ends were phosphorylated and 3' ends dA-tailed, yielding suitable ends for adaptor ligation. Following the incubation at 65 °C, samples were purified with 0.4x AMPure XP beads (section 2.6.10) (24 µL beads added to 60 µL samples). Beads were rinsed with 70% (v/v) ethanol. DNA was eluted from the beads using 61 µL of Milli-Q H_2O at 37 °C for 30-60 min. After elution, the DNA concentration was assessed using the Qubit dsDNA HS kit (section 2.6.7).

Up to 2.5 µg of DNA, depending on how much was eluted from the beads, was carefully transferred to a fresh LoBind tube using a wide-bore pipette tip and the volume was made up

to 60 µL with Milli-Q H_2O where necessary. Adaptor ligation reactions were set up by addition of the following components:

Component	Volume (µL)	Reagent source
Ligation buffer (LNB)	25	ONT (SQK-LSK109 kit)
NEBNext Quick T4 DNA ligase	10	NEB (product number E6057)
Adapter Mix (AMX)	5	ONT (SQK-LSK109 kit)

Tubes were gently flicked to mix and incubated at room temperature for at least 30 min. The samples were purified using 0.4x AMPure beads as above, with the exception of the use of 250 µL of the Long Fragment Buffer (LFB) from the ONT SQK-LSK109 kit for the bead washing steps. After the second wash, DNA was eluted from the beads by addition of 15 µL of Elution Buffer (EB) from the ONT SQK-LSK109 kit and incubation at 30 °C, 300 rpm in a Thermomixer comfort (Eppendorf) for at least 20 min (until beads pelleted as usual in magnetic rack). The concentration of the adaptor-ligated library was assessed with the Qubit dsDNA HS kit (section 2.6.7).

R9.4.1 FLO-MIN106 flow cells were used for each experiment. During the DNA elution from the beads, a flow cell was placed in the MinION to allow equilibration to the run temperature of 34 °C. Runs were set up using the “New Experiment” option in the ONT MinKNOW software. Each flow cell was primed prior to the run. For this, 975 µL of Flush Buffer (FLB) were mixed with 25 µL of Flush Tether (FLT, both from the ONT SQK-LSK109 kit). Air was removed from the priming port with a P1000 pipette before loading 800 µL of the FLB + FLT mix and incubation for approximately 5 min. The remaining 200 µL were then loaded after opening the SpotON sample port cover on the flow cell. Once primed, the flow cell could be loaded with the library. After sufficient incubation to elute the DNA from the beads, 12 µL of the adaptor-ligated sample were transferred to a fresh 1.5 mL LoBind tube and 37.5 µL of Sequencing Buffer (SQB) and 25.5 µL of Loading Beads (LB) from the ONT SQK-LSK109 kit were added. The tube was gently tapped to mix, then half of the mix was carefully loaded drop-wise into the SpotON sample port from a wide-bore pipette tip. The remaining library was gently mixed and loaded. The SpotON sample port cover and priming port were both closed and the MinION lid

placed over the flow cell. The run could then be started from the MinKNOW software. Mux scans were performed automatically to determine the optimal pore usage throughout each run.

2.8 Bioinformatics

All R analysis was performed using `R studio Version 1.3.959`. All IGV visualisations used IGV_2.8.0. Figures were produced using Affinity Designer 1.8.4. Unless otherwise stated, all scripts are available from https://github.com/DNAReplicationLab/EmmaHeronThesis.

2.8.1 BrdU-IP data analysis

Data generated during an Illumina run were downloaded to a departmental server from Illumina Basespace as fastq.gz files. FastQC files were generated to assess the quality of the sequencing data (https://github.com/s-andrews/FastQC). The BrdU-IP data analysis pipeline began with unzipping fastq.gz files to fastq files using the Linux `gunzip` command. Files with the same prefix were concatenated using the Linux `cat` command and piped into the `fastx_barcode_splitter.pl` script, part of the FASTX toolkit (https://github.com/agordon/fastx_toolkit). Commands for this step had the following format (the blue text indicates parts of the commands that varied for each experiment):

```
cat fastq_file_prefix* | fastx_barcode_splitter.pl --bcfile barcode_info.txt --bol --prefix output_file_prefix_
```

This split reads according to barcode information so that each sample could be identified and extracted from the pool of reads. Next, the `fastx_trimmer` (also part of the FASTX toolkit) was used to remove the first five bases from reads (the 4-base barcode and the T complementary to the A-tail added during the library preparation). Files were output from the `fastx_trimmer` as fastq.gz files:

```
sbatch --wrap "fastx_trimmer -f 5 -i input_file_name -o output_file_name_trim.fastq.gz -z"
```

These files were then mapped to the reference genome in 1 kb windows using the `CANmap_IP.sh` script (kindly provided by Dr. C. Müller):

```
perl CANmap_IP.sh -g Spombe -U input_file_name_trim.fastq.gz -s output_file_name
```

This outputted bam alignment files, as well as bed files containing the number of uniquely mapping reads per 1 kb genomic window that could be used for downstream analysis. The *S. pombe* genome reference used was Schizosaccharomyces_pombe_ASM294v2_31. To determine the number of reads mapping to the nuclear genome for each sample, for calculation of contributions to read numbers to the overall pool, `samtools idxstats` (https://github.com/samtools/samtools) was used.

Mapped read counts in bed files were processed to remove rows corresponding to mitochondrial DNA and to re-order the chromosomes to the correct order. These processed bed files were then opened in `R studio` to calculate the Z score of each window and remove windows where the Z score > 2.5. Each window was then normalised to the input to take account of the fraction of the sample in the pool. The resulting data were used for plotting the number of IP reads across each chromosome using `R`'s `ggplot2` package.

To detect peaks of BrdU incorporation in the IP data, the Z score-filtered and input-normalised bed files first had to be converted to wig format using the `bed_to_wig_no_rounding_5_column.pl` script (kindly provided by Dr. C. Müller):

```
perl bed_to_wig_no_rounding_5_column.pl -in input_file_name.bed -out output_file_name.wig
```

The script used for peak detection (`peakdetect.py`, kindly provided by Dr. C. Müller) required an additional chromosome at the end of the input wig file. Thus, an extra chromosome called chrZZZ was created, that consisted of the first 1000 lines from one of the sample wig files (these lines were deleted from each file in a later step):

```
echo 'variableStep chrom=chrZZZ' >> file_name.wig
grep -v chr wig_file_to_copy_lines_from.wig | head -1000 >> output_file_name.wig
```

The resulting .wig files could be analysed for peaks using `peakdetect.py`, with the peak threshold given for each analysis:

```
python peakdetect.py input_file_name.wig threshold_for_calling_a_peak > output_file_name_peaks.wig
```

The wig file output by the `peakdetect.py` script was then converted back to bed format using the `wig_to_bed.pl` script (kindly provided by Dr. C. Müller):

```
perl wig_to_bed.pl --in input_file_name_peaks.wig --bed output_file_name_peaks.bed
```

The chrZZZ chromosome added to the end of the file was removed, and the resulting file was copied with the file name changing to having peak_locations_ as the prefix:

```
grep -v chrZZZ file_name_peaks.bed > file_name_peaks_no_chrZZZ.bed
cp file_name_peaks_no_chrZZZ.bed peak_locations_file_name.bed
```

The number of lines in each bed file was determined using the Linux `wc` command (used to ascertain the number of peaks detected):

```
wc -l peak_locations_file_name.bed
```

For comparison with the published PU-seq origin dataset (Daigaku et al. 2015) the `closestBed bedtools` command (https://bedtools.readthedocs.io/en/latest/) was used:

```
closestBed -a peak_locations_file_name.bed -b PU-seq_oris.bed > closest_a_to_b.txt
```

To plot the distribution of the distances obtained from the `closestBed` command, the `bed_comparison.r` script was used (kindly provided by Dr. C. Müller):

```
Rscript bed_comparison.r closest_a_to_b.txt
```

To generate random genomic windows for comparison with the PU-seq origins, the `bedtools randomBed` command (https://bedtools.readthedocs.io/en/latest/) was used, with the length (`-l`) of each window and the number of windows to produce (`-n`) specified:

```
randomBed -l window_length -n number_of_windows -g pombe_chr_sizes.txt
> random.bed
```

The resulting random.bed file was processed in `R studio` to remove unnecessary columns and to re-order the rows by chromosome number and coordinate. The modified random.bed file was exported from `R studio` and compared to the PU-seq origins using the `closestBed` command, as above.

2.8.2 D-NAscent data analysis

Raw nanopore signal data generated on each MinION flow cell were saved as FAST5 files by ONT's `MinKNOW` software. ONT's `Guppy` software (version 3.6) was then used for basecalling, thereby converting the signal information to sequence information in FASTQ format. This was kindly performed by Dr. Carolin Müller, using the following commands (the blue text indicates parts of the commands that varied for each experiment):

```
sudo mkdir experiment_name_basecalled_Guppy3.6
sudo /path/to/software/ont-guppy/bin/guppy_basecaller -i
experimental_data_directory/ -s experiment_name_basecalled_Guppy3.6/
-c /path/to/software/ont-guppy/data/dna_r9.4.1_450bps_fast.cfg -r -x
'cuda:0'
```

Next, it was necessary to concatenate the resulting FASTQ files to produce a single reads.fastq file for subsequent analysis. This was performed using the Linux `cat` command in a working directory unique to each experiment as follows:

```
cat /path/to/experimental/data/experiment_name_basecalled_Guppy3.6/*
.fastq > reads.fastq
```

The concatenated FASTQ data were then aligned to the reference *S. pombe* genome sequence using `minimap2`, version 2.17 (https://github.com/lh3/minimap2). The output was directly piped to `samtools` (https://github.com/samtools/samtools) for conversion from a SAM to a BAM file format (`view`) and sorting aligned sequences with respect to genome position (`sort`):

```
/path/to/software/minimap2-2.17/minimap2 -ax map-ont -t 50
Schizosaccharomyces_pombe.ASM294v2.31.dna.genome.fa reads.fastq |
samtools view -Sb - | samtools sort -o alignments.sorted.bam
```

The alignments.sorted.bam file was then indexed using `samtools index`:

```
samtools index alignments.sorted.bam
```

Subsequent data analysis and processing were performed using `DNAscent` v2 software (Boemo 2020) (https://github.com/MBoemo/DNAscent). For this, the following commands were necessary to direct the GPU to required software libraries:

```
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/usr/local/cuda-10.1/lib64
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/usr/local/cuda/lib64
```

Next, a link between the ONT molecule read IDs and the raw signal data in the FAST5 files was created to enable `DNAscent detect` to locate the signal information for each read in the bam file. This was achieved using the `DNAscent index` sub-programme:

```
/path/to/DNAscent_v2/software/directory index -f
/path/to/experimental/data/experiment_name -s
/path/to/experimental/data//experiment_name_basecalled_Guppy3.6/sequ
encing_summary.txt
```

Next, `DNAscent detect` was run to determine the probability of each thymidine in a read being a BrdU. The `-l 30000` argument specified that the minimum read length to be included was 30 kb:

```
/path/to/DNAscent_v2/software/directory detect -l 30000 -b
alignments.sorted.bam -r
Schizosaccharomyces_pombe.ASM294v2.31.dna.genome.fa -i index.dnascent
-o output_30kb_v2.detect -t 25 --GPU 0
```

To predict the probability of a leftward- or rightward-moving fork at each thymidine based on the BrdU assignments by `detect`, the `forkSense` sub-programme was used. The optional `--markOrigins` argument was included to derive predicted initiation sites based on the fork direction probabilities:

```
/path/to/DNAscent_v2/software/directory forkSense -d
output_30kb_v2.detect -o forkSense_Oris.forkSense --markOrigins -t 10
```

To produce the `detect` ensemble plots, a threshold of 0.7 was selected for the probability of each thymidine being a BrdU to assign each position as a BrdU. The fraction of thymidines in each genomic window being called as BrdU based on this threshold was then determined using the `plotDetect_out_text_chrX.py` scripts (X = 1, 2, or 3, depending on which chromosome was being analysed):

```
python plotDetect_out_text_chrX.py output_30kb_v2.detect
```

The output was plotted using `R`'s `ggplot2` package, along with the Z score-filtered IP data from the same DNA sample.

To convert `detect` and `forkSense` information to bedgraphs for visualisation in IGV's genome browser, the `dnascent2bedgraph.py` script was used:

```
python dnascent2bedgraph.py -d output_30kb_v2.detect -f forkSense_Oris.forkSense -minLength 30000 -o bedgraphs
```

For the visualisation of the genome-wide forkSense origin predictions, the origins_DNAscent_forkSense.bed files were visualised in IGV along with the Z score-normalised IP data from the same DNA sample, the PU-seq-called origins (Daigaku et al. 2015), sort-seq data and the relative DNA content of the sample harvested 75 min after elutriation, normalised to the sample harvested 45 min after elutriation (Daigaku et al. 2015). Both the sort-seq and mid-S phase time point datasets were kindly provided by Dr. Carolin Müller.

2.8.3 BNG data analysis

Irys data were analysed using BNG's IrysView software to generate *de novo* genome assemblies. *In silico* digestion of the sacCer3 and Schizosaccharomyces_pombe_ASM294v2_31 references was performed using BNG's Knickers and LabelDensityCalculator programmes. The ability to produce reference cmaps by *in silico* digestion has since been moved to Bionano Access, which was used for the DLE-1 analysis of sacCer3.

Saphyr raw molecules were aligned to a reference DLE-1 cmap file using the `pipelineCL.py` script in `BionanoSolve` (with help from Y. Chang from BNG) (https://bionanogenomics.com/support/software-downloads/#bionanoaccess):

```
python2.7 12162019//pipelineCL.py -y -x -z -T 48 -j 5 -b
/path/to/data/RawMolecules_sample.bnx -l /output/directory -e
output_file_prefix -a
/path/to/BionanoSolve/Solve3.5.1_01142020/RefAligner/1.0/optArgument
s_haplotype_saphyr.xml -t
/path/toBionanoSolve/Solve3.5.1_01142020/RefAligner/1.0/ -r
/path/to/reference/cmap/sacser3_DLE-1_0kb_0labels.cmap -F 0
```

This output bnx, xmap and cmap files for downstream analysis. The bnx file provides information about the positions of each label in each molecule and the associated quality scores. The xmap shows the alignment information for each molecule in relation to the reference, giving the genomic location of the molecules. The cmap details the label information of each molecule, listing the positions of the detected labels with regard to the start of the molecule, and the GmeanSNR of each signal (https://bionanogenomics.com/support-page/data-analysis-documentation/).

To combine the alignment information from the xmap and cmap files, a custom `R` script was kindly written by Ms. Phoebe Oldach, which was subsequently adapted to allow the calculation of the genomic coordinate of each detected label. In addition, a custom `R` script written by BNG (`2_color_analysis_for_customer.R`) was kindly shared by Dr. Simon McGowan and was used to obtain label channel 2 (LC2) genomic location coordinates. This output a .txt file with information about the molecule identity (ID), the position of each LC2 label within a molecule, the chromosome to which the molecule mapped, the position on the chromosome to which each LC2 signal aligned, and the orientation of a molecule relative to the reference. Single molecule label information output by each of these scripts was visualised using IGV, along with confirmed and likely *S. cerevisiae* origins of replication from OriDB (Nieduszynski et al. 2007; Siow et al. 2012).

To compare the LC2 information with previously-published replication timing data (Müller et al. 2014), the GSM1180749_T9475_Illumina_normalised dataset from this publication was opened in `R studio` and the rows were re-ordered according to the relative DNA copy number (from lowest to highest). The number of rows was found to be 11,469 by the `nrow()` command. The re-ordered data were output as a bed file and divided into ten equally-sized bins (1147 rows, or 1146 for the final bin) using the Linux `split` command:

```
split -l 1147 T9475_sortseq_sort_by_RT.bed T9475_sortseq_sort_by_RT.bed.
```

This output files suffixed with .aa, .ab, .ac... These were changed to keep .bed as the suffix by altering .aa to T9475_sortseq_sort_by_RT1.bed, for example. The ten bed files were then opened in `R studio` and the rows in each file were re-ordered by chromosome and location, and converted to 1 kb windows. The resulting data frames were output as bed files, and the intersect between the locations of all LC1 or LC2 signals from the *pGAL-Os.TIR1, CDC9-IAA17* aligned Saphyr dataset was determined using `bedtools intersect` (https://bedtools.readthedocs.io/en/latest/):

```
intersectBed -a LC2_locations.bed -b T9475_sortseq_sort_by_RT_1_ordered.bed > LC2_intersect_RT1.txt
```

The number of LC1 or LC2 signals that mapped to each bin was then determined using the Linux `wc` command:

```
wc -l LC2_intersect_RT*
```

These values were read into `R studio` as a data frame and plotted as a bar graph using `ggplot2`.

To trial calling LC2 signals as tracts of OF labels, a modified version of the `direction_bed_seg_new_format.py` script (Klein et al. 2017) (https://github.com/tborrman/DNA-rep/tree/master/BioNano) was used. For the different

parameters, the values in lines 27 (`if total_neighbors > x`), 25 (`if abs(test_label - query_label) < x`), 34 (`if abs(red_labels[neighbor_idx] - red_labels[keep_idx]) < x`), 71 (`if abs(current_pos - next_pos) < x and begin_segment:`) and 74 (`elif abs(current_pos - next_pos) >= x and not begin_segment:`) were altered.

To calculate the cumulative sum of LC2 GmeanSNR values in single molecules, the data output by P. Oldach's script were opened as a data frame in `R studio` and, using the `dplyr` package, a column called cumulativeScore was added by the `mutate` command which calculated the cumulative sum of signal for each molecule using the `group_by` and `cumsum` commands. The data frame was then filtered to extract individual molecules for plotting using `ggplot2`.

Chapter 3: Applying D-NAscent to the study genome replication in *S. pombe*

3.1 Introduction

A recent study from the research group described a novel method for detecting nucleotide analogue signal currents on extremely long nanopore traces (D-NAscent) in *S. cerevisiae* (Müller et al. 2019). This was used to analyse replication dynamics in long single molecules of known sequence genome-wide by identifying regions of BrdU incorporation during replication by nanopore sequencing. The aim of this project was to develop a protocol for D-NAscent in *S. pombe*, to enable the single molecule analysis of replication dynamics in a model system with a genome organisation more analogous to that of metazoan genomes, but with the ease of use and manipulation of yeast (Hoffman, Wood, and Fantes 2015). To apply D-NAscent to the study of genome replication in *S. pombe*, a number of experimental requirements had to be met (Figure 3.1). To maximise the number of molecules undergoing replication during the BrdU pulse, and therefore to increase the proportion of BrdU-labelled molecules analysed on a MinION sequencing run, it was necessary to synchronise cell cultures. In addition, since the *S. pombe* genome does not encode the proteins required for thymidine salvage from the medium (Sivakumar et al. 2004), it was necessary to use a strain that had been modified to stably express genes encoding an equilibrative nucleoside transporter (ENT) and a thymidine



Figure 3.1*: Schematic showing the experimental requirements of D-NAscent. To apply D-NAscent to the study of genome replication in* S. pombe*, it was necessary to achieve cell cycle synchronisation, use cells that were modified to enable the genomic incorporation of BrdU added to the medium, and to extract high molecular weight (HMW) molecules for nanopore sequencing.*

kinase (TK). Finally, to fully exploit the potential of the nanopore sequencing platform, it was important to optimise DNA extraction to obtain high molecular weight molecules.

3.2 Verification of G1 cell cycle synchronisation in *S. pombe*

To address the aim of cell cycle synchronisation, the reversible G1 arrest of a *cdc10-m17* strain (that also expressed the human ENT (hENT) and the *Herpes simplex* virus TK (HSV-TK)), kindly provided by collaborators in the Carr group (University of Sussex), was validated. This allele encodes a temperature-sensitive (ts) version of Cdc10, a subunit of the MluI cell cycle box- (MCB)-binding factor (MBF) transcriptional regulation complex (Bähler 2005). This complex is required for the activation of gene expression in the transition from G1 to S phase. Incubation of the *cdc10-m17* strain at its restrictive temperature results in the accumulation of cells in G1 phase, with a DNA content of 1C (Carlson et al. 1997; Nurse, Thuriaux, and Nasmyth 1976). Therefore, a preliminary goal was to reproduce the G1 phase arrest by incubation at the restrictive temperature, and to confirm synchronous S phase progression following release at the permissive temperature. A culture was grown at the permissive temperature (25 °C) before transfer to the restrictive temperature (36 °C) for 3 h 40 min. After this time, the culture was rapidly cooled back to the permissive temperature and incubated for a further 2 h. Samples were taken for flow cytometry analysis of DNA content throughout (Figure 3.2). This showed an accumulation of cells with a DNA content of 1C after incubation at the restrictive temperature when compared to the asynchronous sample, indicative of an arrest in G1 phase. By 60 min after the release at the permissive temperature, the majority of cells had a DNA content between 1C and 2C, consistent with ongoing genome replication. After 90 min at the permissive temperature, the majority of cells in the population had reached a DNA content of 2C, suggesting completion of genome replication. Therefore, it was possible to reproduce a G1 arrest and synchronous S phase progression using the *cdc10-m17* strain and to monitor this by flow cytometric analysis of DNA content.

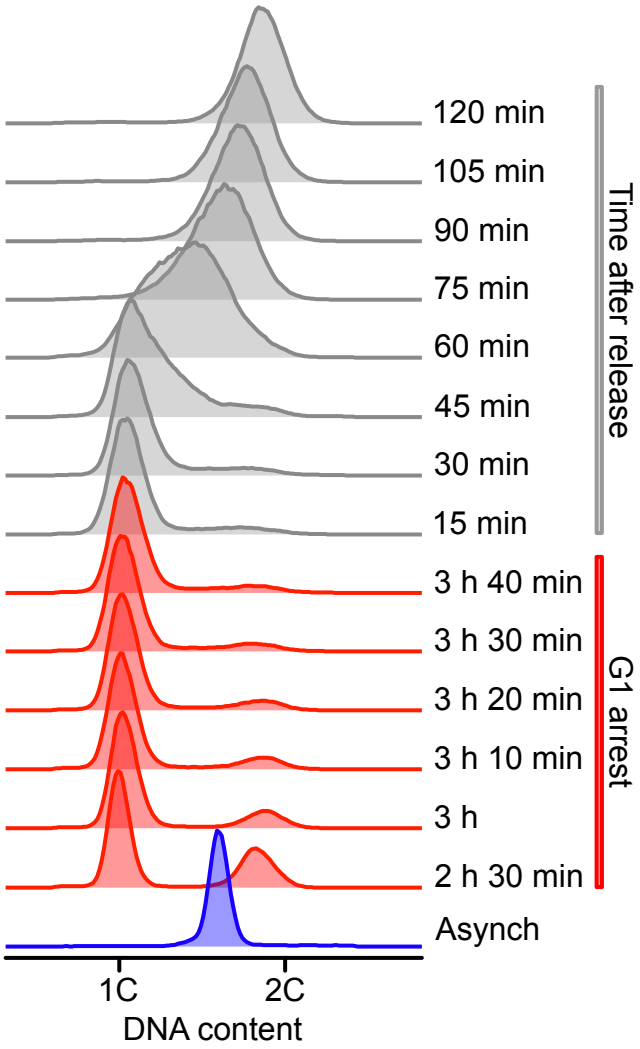


***Figure 3.2**: Histograms of DNA content during a G1 arrest and release experiment in a* cdc10-m17 *mutant. A* cdc10-m17 *ts mutant was grown to early exponential phase at its permissive temperature (25 °C) before transfer to the restrictive temperature (36 °C) for a total of 3 h 40 min. Samples were collected for flow cytometry analysis at the indicated time points. After 3 h 40 min at 36 °C, the culture was cooled to 25 °C and incubated at this temperature for 2 h, with samples collected for flow cytometry every 15 min. Incubation at 36 °C resulted in the majority of the population having a DNA content of 1C after 3 h 30 min – 3 h 40 min. Following the transfer back to the restrictive temperature, most cells had reached a DNA content of approximately 2C by 90 min after the G1 release.*

Next, a G1 arrest/release experiment was performed with a prolonged incubation at the restrictive temperature to identify the optimal conditions for maximal culture arrest in G1 phase while avoiding escape from the arrest and leakage into S phase. Based on the result in Figure

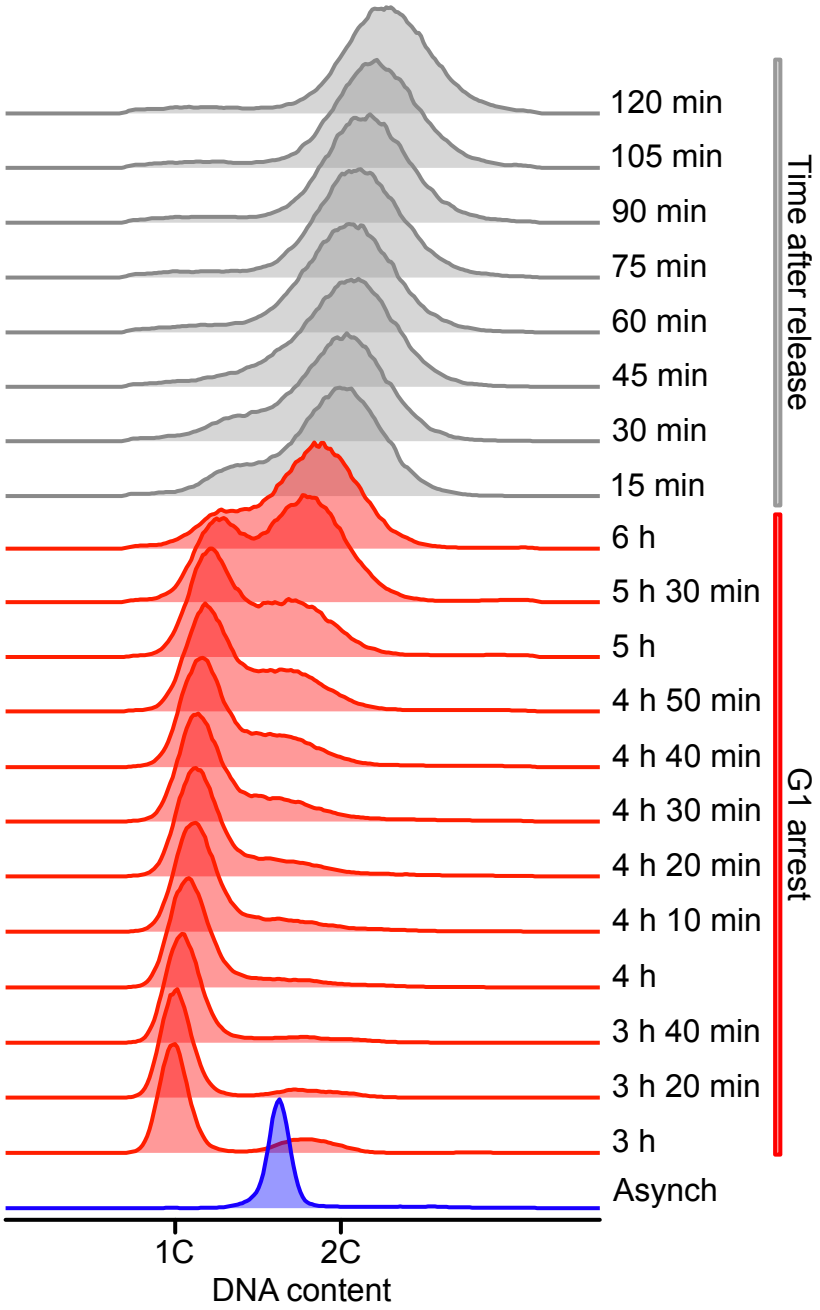


***Figure 3.3**: Histograms of DNA content during an extended G1 arrest and release experiment in a* cdc10-m17 *mutant. As in Figure 3.2, with the exception of a prolonged incubation at 36 °C. After 4 h at 36 °C, the DNA content of a subset of the population began to increase above 1C. The proportion of cells that had this increased DNA content continued to rise as the duration of the incubation at 36 °C increased.*

3.2 and personal communication with Dr. Y. Daigaku (Tohoku University), the optimal time for this was predicted to be approximately 3 h 30 min – 3 h 40 min. However, to assess the time at which cells began to escape the arrest, the culture was incubated for a total of 6 h at the

restrictive temperature. The cells were then incubated at the permissive temperature for a further 2 h and the total DNA content was assessed throughout by flow cytometry (Figure 3.3). By 4 h at the restrictive temperature, a small fraction of the population had a DNA content greater than 1C, indicating that these cells had escaped the G1 arrest and had begun genome replication. This was consistent with the prediction that the cells would begin to leak into S phase after a prolonged incubation at the restrictive temperature. The proportion of cells that escaped the arrest increased as the incubation time at the restrictive temperature increased. Release at the permissive temperature followed the trend in the population's cell cycle progression, as remaining cells with a DNA content of approximately 1C appeared to proceed to replicate their genomes such that the majority of the population had a DNA content centring around 2C by 45 min at the permissive temperature.

In summary, it was concluded that the optimal length of incubation at the restrictive temperature for this *cdc10-m17* strain was approximately 3 h 30 min – 3 h 40 min; a longer incubation time was confirmed to allow an escape from the arrest.

3.3 Assessment of BrdU incorporation by immunoprecipitation followed by short-read sequencing

Since a protocol for cell cycle synchronisation in G1 phase and the subsequent S phase had been validated, it was then necessary to confirm BrdU incorporation during genome replication in the *cdc10-m17, adh:hENT1, adh:hsv-TK* strain. As a read-out for BrdU incorporation, immunoprecipitation (IP) was used to enrich for BrdU-containing molecules, which were then analysed by Illumina sequencing, as this is an established method for studying genome replication in both *S. cerevisiae* (Haye-Bertolozzi and Aparicio 2018) and *S. pombe* (Daigaku et al. 2017). Published studies involving BrdU-IP from *S. pombe* had used from 0.5 µM BrdU (Daigaku et al. 2017) up to approximately 326 µM BrdU (100 µg/mL) (Li, Green, and Forsburg 2013; Li et al. 2011). In this work, the lower extreme of this range was investigated, since it had previously been shown that 0.5 µM BrdU for a single S phase gave equivalent signal as higher concentrations when detecting labelled DNA by immunofluorescence, and that much

higher concentrations were detrimental to cell viability and cell cycle progression (Sivakumar et al. 2004). In addition, previously published single-cell and single-molecule studies had used 0.5 μM BrdU to label DNA for epifluorescence microscopy (Sivakumar et al. 2004) to 2 μM BrdU to label DNA for combing (Kaykov et al. 2016; Kaykov and Nurse 2015), supporting the evidence that the lower concentrations were sufficient. However, an important difference between all of these approaches and D-NAscent is the enrichment of BrdU-containing molecules by IP or amplification of the BrdU signal by antibody detection, which are absent in D-NAscent. Therefore, it was possible that a concentration that was optimal for BrdU-IP might not be appropriate for D-NAscent. Initially, incorporation was assessed by BrdU-IP, which would subsequently be used for validation of D-NAscent data (section 3.5), to identify potential experimental conditions that might have been suitable.

3.3.1 Addition of 1 μM BrdU during a G1 arrest

Based on the evidence from the literature discussed above, a concentration of 1 μM was chosen for the first BrdU-IP-seq experiment. To label the earliest replicating regions of the genome, BrdU was added to the culture during a G1 arrest to allow time for BrdU uptake from the medium and intracellular processing into a substrate for the replicative polymerases. After allowing 20 min for this uptake and processing, cells were released from the G1 arrest by cooling to the permissive temperature. Two control samples were harvested, one at 3 h 15 min into the G1 arrest, immediately before the addition of BrdU, and the other at 3 h 35 min, immediately before the release from G1. These were to provide negative controls for the BrdU-IP as neither were expected to have BrdU-containing genomic DNA, the former because it had not yet been exposed to BrdU, and the latter because it had not begun nuclear DNA replication (according to the bulk DNA content, Figure 3.4). Given that it had previously been shown that the majority of cells had reached a DNA content of 2C and therefore replicated their genome by 90 min after the release (Figure 3.2), the final sample was harvested at this time point. The G1 arrest and subsequent release into S phase was monitored by flow cytometry of total DNA content (Figure 3.4). This indicated that the majority of cells had been arrested in G1 phase by 3 h 35 min at the restrictive temperature. Following release at the permissive temperature,

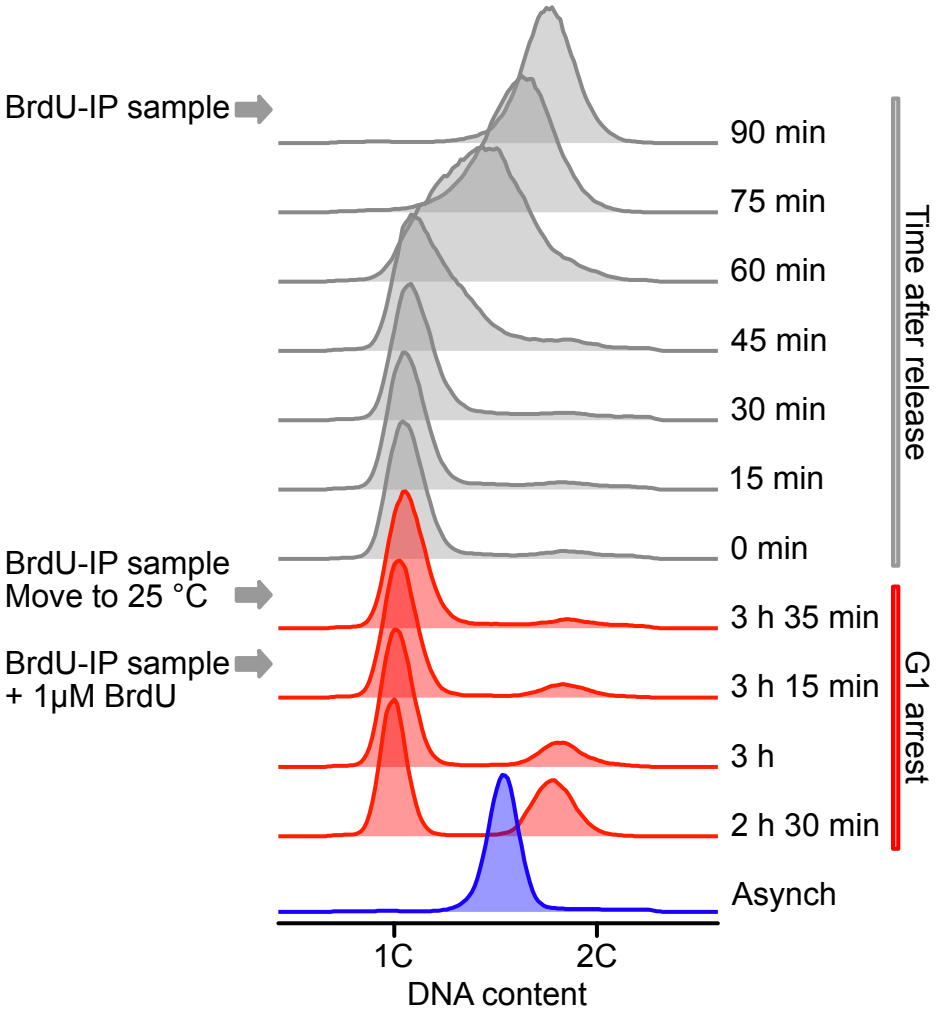


***Figure 3.4**: Histograms of DNA content during a G1 arrest and release experiment in a* cdc10-m17, adh:hENT1, adh:hsv-TK *strain with 1 µM BrdU addition during the arrest. Cells were arrested with a DNA content of 1C by incubation at the restrictive temperature. After 3 h 15 min at the restrictive temperature, a control sample was harvested for BrdU-IP-seq and 1 µM BrdU was added to the remaining culture. After a further 20 min at the restrictive temperature, another control sample was harvested for BrdU-IP-seq and the remaining cells were shifted to the permissive temperature. By 90 min after the release, the majority of cells had a DNA content of 2C. The final BrdU-IP-seq sample was harvested at this time point.*

progression through S phase appeared to take place with a good level of synchrony, with the majority of cells reaching a DNA content of 2C by 90 min following release from the G1 arrest, as in Figure 3.2. This indicated that the addition of 1 µM BrdU during the G1 arrest did not perturb bulk replication dynamics during S phase.

To assess whether BrdU had been incorporated at the earliest replicating regions of the genome, as anticipated if the BrdU concentration had been limiting, DNA was extracted from samples harvested at the indicated time points (Figure 3.4) and subjected to BrdU-IP. Prior to the IP step, samples from the different time points were barcoded and pooled, to allow direct inter-sample comparisons. Libraries were then analysed by Illumina sequencing to determine the number of reads corresponding to each sample (Table 3.1). To account for differences in pooling, indicated by the fraction of reads from each barcoded sample making up the input pool, the reads for each IP sample were normalised to their corresponding input sample (final column in Table 3.1). This showed that > 95% of the IP reads were derived from the sample harvested 90 min after release, consistent with the expectation that the other two samples

Table 3.1*: Summary of read counts for Illumina libraries. Barcoded input and IP samples were pooled and a unique index was added during amplification of each pool. The barcodes added to each sample prior to pooling were used to assess the number of reads for each sample within each pool. Pooling was performed such that this should have been equal for each sample. The fraction of each sample in the IP pool was normalised to the corresponding input to account for differences in the sample pooling prior to IP.*

Index pool	Sample	Read count	Fraction of reads in pool	Input-normalised fraction of reads in pool
Input	3 h 15 min in G1 arrest	183783	0.500	NA
	3 h 35 min in G1 arrest	70906	0.193	NA
	90 min after release	112543	0.306	NA
IP	3 h 15 min in G1 arrest	3131	0.006	0.003
	3 h 35 min in G1 arrest	14859	0.027	0.042
	90 min after release	538509	0.968	0.955

should have had either no or minimal levels of BrdU incorporation (for the 3 h 15 min and 3 h 35 min samples, respectively). The contribution of approximately 4% of reads to the IP pool from the 3 h 35 min sample suggested that there had been limited BrdU incorporation by this time point, consistent with a small proportion of the cell population escaping the arrest and beginning genome replication.

To assess the pattern of BrdU incorporation in the IP samples, the number of reads mapping to each 5 kb genome window was normalised based on the fraction of reads each barcode contributed to the input pool. These were plotted across the genome to determine the pattern of BrdU incorporation in the three chromosomes (Figure 3.5). This showed that there was minimal BrdU signal for the samples harvested after 3 h 15 min and 3 h 35 min at the restrictive temperature, as expected. The sample harvested 90 min after the release from G1 showed widespread BrdU incorporation across the genome. However, a clear pattern of localised incorporation around the earliest firing initiation sites could not be discerned from these data, due to the relatively uniform incorporation. To aid visualisation, a 500 kb window on chromosome III was plotted (lowest plot). Clearly defined peaks of BrdU incorporation, for example, associated with known replication origins, were not observable in these data. Therefore, it was confirmed that the strain incorporated BrdU specifically during replication. However, the presence of BrdU throughout S phase gave incorporation across the genome and so this experimental setup was not suitable for the detection of early replicating regions.

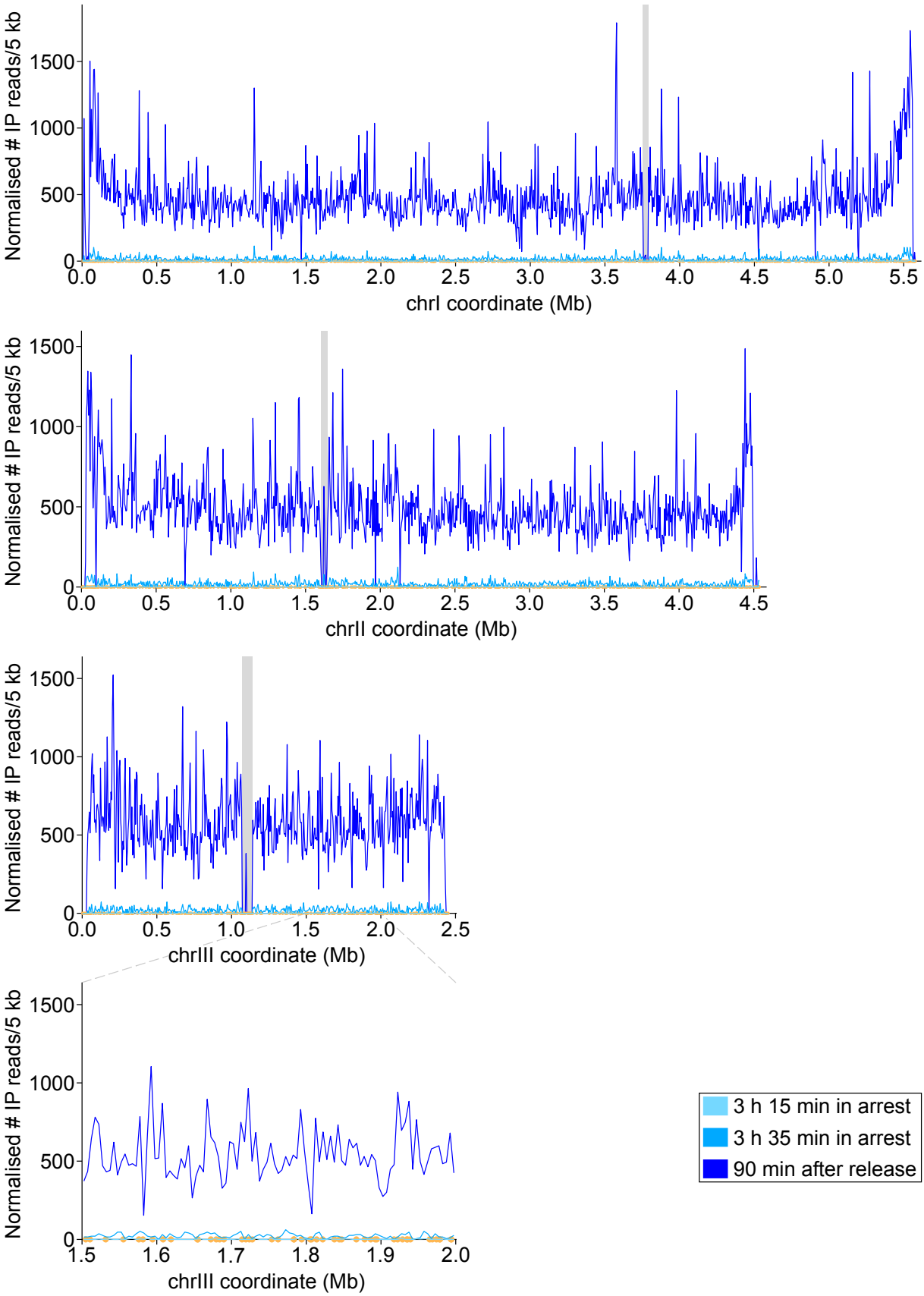


***Figure 3.5**: Number of BrdU-IP reads mapping to 5 kb bins across the genome. Orange circles along the x-axis mark sites designated as origins of replication by polymerase-usage sequencing (PU-seq)* (Daigaku et al. 2015). *Vertical grey bars mark the centromeres. Windows*

were filtered to remove signal spikes by removing those bins with a Z score > 2.5 and each dataset was normalised to account for differences in pooling (Table 3.1). All three chromosomes were plotted, with a zoomed in plot of 500 kb on chromosome III to aid visualisation.

3.3.2 Addition of 0.5 μM BrdU at different time points during G1 arrest and release

To assess the impact of changing the time of the BrdU pulse, a cell cycle synchronisation experiment was performed with BrdU additions at different time points during the arrest and release. Based on the population replication timing data, it was predicted that BrdU addition at specific times during S phase might result in BrdU incorporation at different genomic locations. For this experiment, the concentration of BrdU was reduced to 0.5 μM to try and limit incorporation to regions replicated shortly after the pulse. The starting culture was split before the first BrdU addition, and BrdU was added to the different culture aliquots at various time points throughout S phase, as described below (Figure 3.6). In this experiment, all samples came from the same starting culture that was arrested in G1 phase by incubation at the restrictive temperature. Immediately prior to the first BrdU addition, the culture was split into two flasks: one of which had BrdU added during the G1 arrest, 20 min before the release into S phase (as in section 3.3.1), and one which did not have BrdU added at this time. After a total of 3 h 35 min at the restrictive temperature (20 min after the culture was first split), both flasks were shifted to the permissive temperature to release cells into S phase. The culture which had not yet been exposed to BrdU was then divided equally into five flasks. BrdU was added to one flask at the point of release (0 min), while the four remaining flasks had a BrdU addition at 15, 30, 45, or 60 min after release from the G1 arrest. Cell cycle progression through S phase was very similar for the five cultures, as determined by flow cytometry analysis of total DNA content, with the majority of the population reaching a DNA content of 2C by 90 min after the release (Figure 3.6).

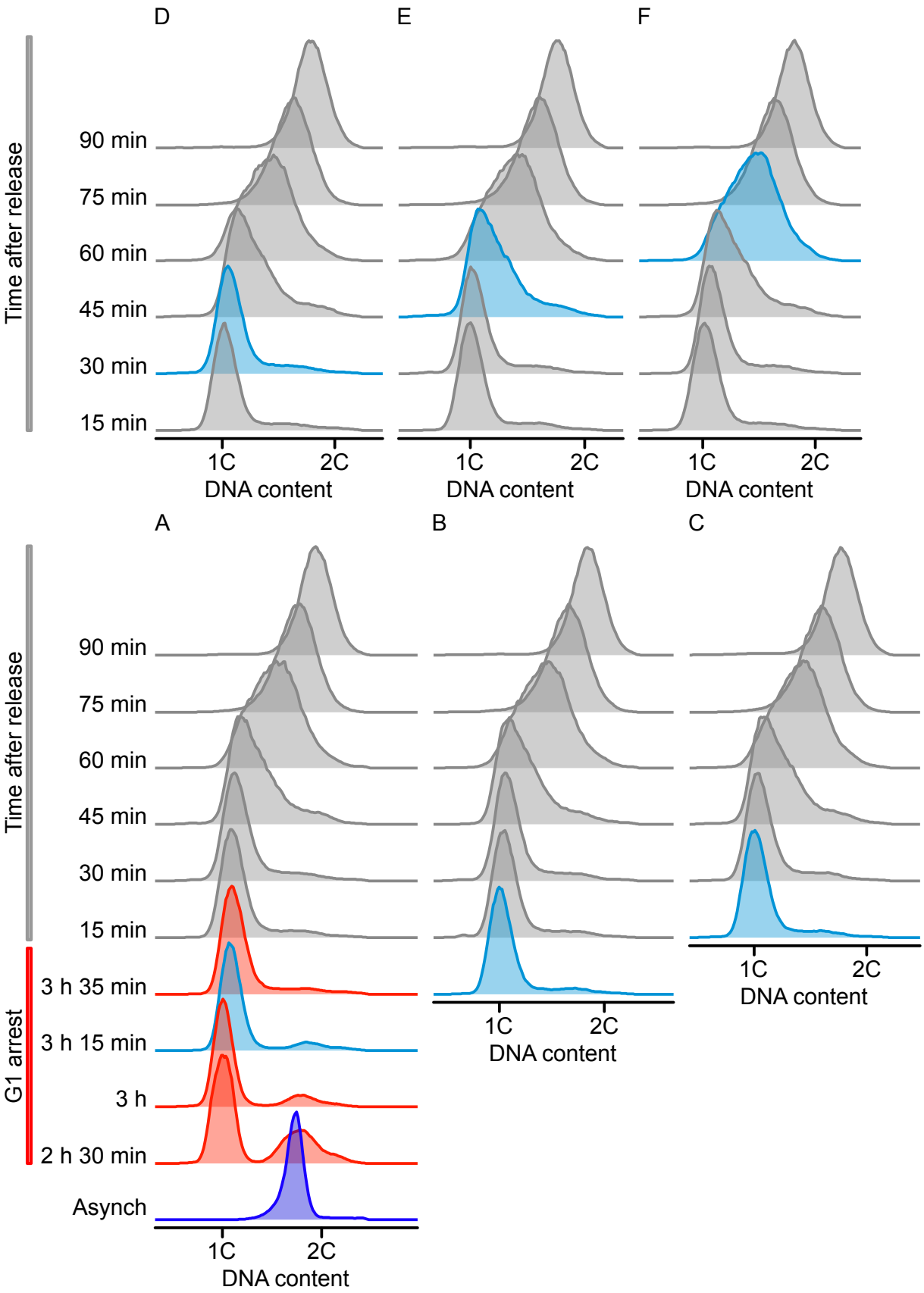


***Figure 3.6**: Histograms of DNA content during a G1 arrest and release experiment in a* cdc10-m17, adh:hENT1, adh:hsv-TK *strain with 0.5 μM BrdU addition at different time points during the experiment. In each case, the time point at which BrdU was added to a sub-culture is*

*highlighted by the light blue profile. (**A**) An initial culture was incubated at the restrictive temperature to trigger a G1 arrest. After 3 h 15 min at the restrictive temperature, a control sample for BrdU-IP-seq was harvested, and an aliquot of the culture was transferred to a separate flask, and supplemented with BrdU (final concentration of 0.5 μM). After a total of 3 h 35 min at the restrictive temperature, another control sample for BrdU-IP-seq was harvested from the culture that had been exposed to BrdU, then the two flasks were both rapidly cooled to the permissive temperature. The culture that had not yet been treated with BrdU was divided into five separate cultures at this point, one of which (**B**) had 0.5 μM BrdU added immediately (0 min after the release). BrdU was added to a final concentration of 0.5 μM to the remaining sub-cultures at 15 min (**C**), 30 min (**D**), 45 min (**E**), or 60 min (**F**) after the shift to the permissive temperature. Flow cytometry samples were taken at the time points shown. Samples for DNA extraction and BrdU-IP were taken from each sub-culture 90 min after the shift to the permissive temperature.*

DNA was purified from the two samples harvested at 3 h 15 min and 3 h 35 min into the G1 arrest and from each of the samples taken 90 min post-release, and processed for BrdU-IP and Illumina sequencing. The samples were barcoded and pooled prior to the IP to allow direct comparison between the different sub-cultures. Again, the fraction of reads in the IP pool was normalised to account for differences in pooling as indicated by the input samples (Table 3.2). Interestingly, this showed that approximately 82% of the reads in the IP pool originated from the sample to which BrdU had been added 20 min before release from the G1 arrest and which was harvested 90 min after the release (BrdU at -20 min). Even the sample which had BrdU added at the time of the release contributed < 6% of the reads in the IP pool, with the contribution of each sample decreasing as the BrdU addition had become later. This suggested that, under these experimental conditions, BrdU added when cells were beginning or had already begun DNA replication was not incorporated by the replication machinery. This could have been due to increasing endogenous thymidine synthesis (Håkansson et al. 2006) which might have inhibited BrdU uptake and/or processing, but this was not investigated further.

***Table 3.2**: Summary of read counts for Illumina libraries based on unique barcodes following sample splitting. Samples shaded in grey were harvested 90 min after the release from G1, with BrdU added to the culture at the indicated time points relative to the release (e.g. BrdU at -20 min had BrdU added 20 min prior to the release into S phase). The fraction of each sample in the IP pool was normalised to the corresponding input to account for differences in the sample pooling prior to IP.*

Index pool	Sample	Read count	Fraction of reads in pool	Input-normalised fraction of reads in pool
Input	3 h 15 min in G1 arrest	98671	0.137	NA
	3 h 35 min in G1 arrest	55962	0.078	NA
	BrdU at -20 min	89644	0.124	NA
	BrdU at 0 min	72711	0.101	NA
	BrdU at 15 min	41203	0.057	NA
	BrdU at 30 min	116923	0.162	NA
	BrdU at 45 min	134471	0.187	NA
	BrdU at 60 min	110463	0.153	NA
IP	3 h 15 min in G1 arrest	8395	0.014	0.012
	3 h 35 min in G1 arrest	12903	0.021	0.033
	BrdU at -20 min	508033	0.846	0.823
	BrdU at 0 min	27961	0.047	0.056
	BrdU at 15 min	10419	0.017	0.037
	BrdU at 30 min	12783	0.021	0.016
	BrdU at 45 min	14562	0.024	0.016
	BrdU at 60 min	5692	0.009	0.007

To assess the distribution of BrdU signal in the different samples, the normalised number of reads mapping to each 5 kb genomic window was plotted across the genome, as in Figure 3.5. Consistent with the proportion of reads from each sample following IP, most of the samples showed very low levels of BrdU signal across the chromosome (Figure 3.7). The only sample to show substantial BrdU signal was that which had received BrdU 20 min before release from the G1 arrest and which was harvested 90 min after the release. However, as in Figure 3.5, this sample showed high levels of noise and signal spikes along the chromosomes, but no

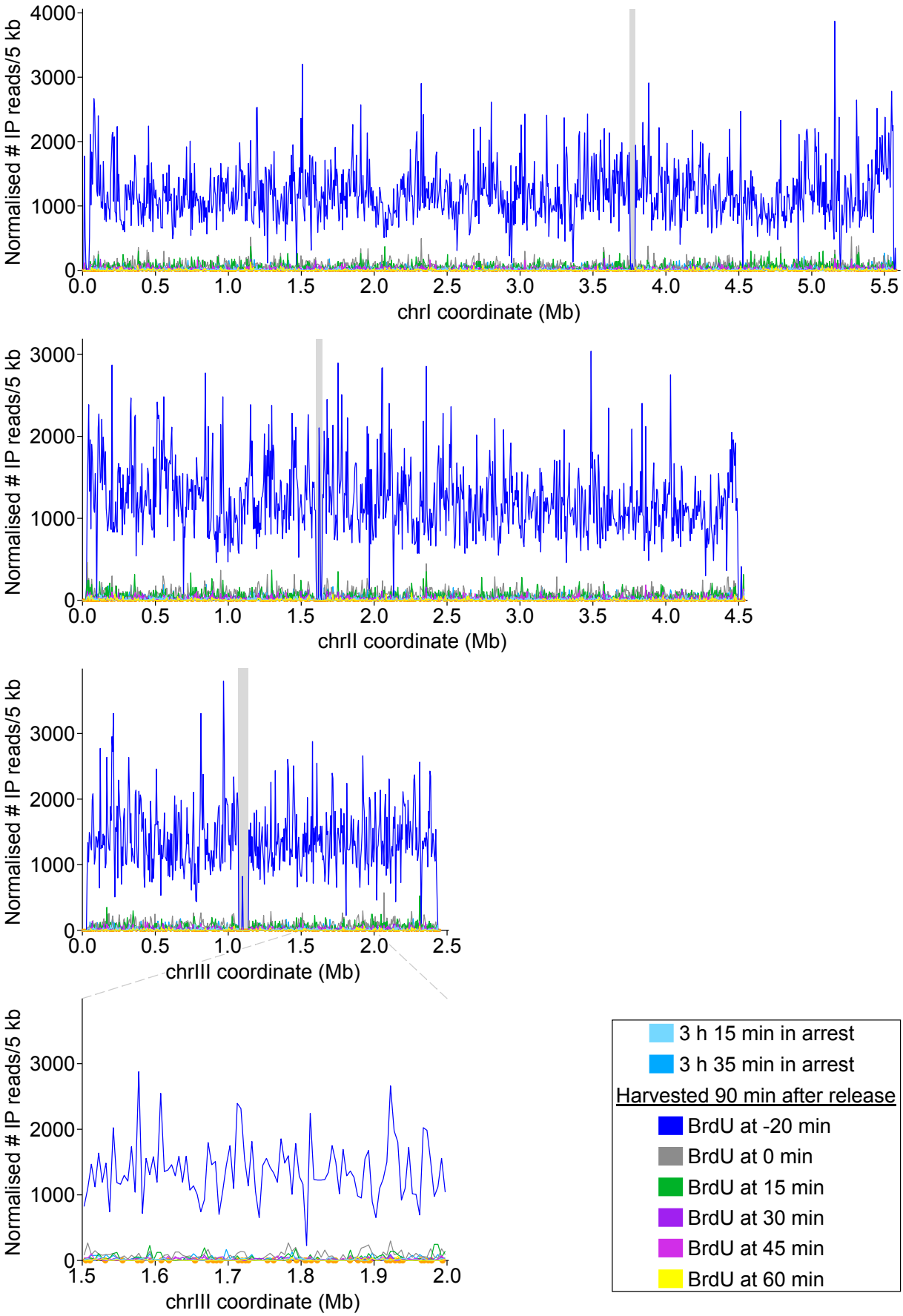


***Figure 3.7**: Number of BrdU-IP reads mapping to 5 kb bins across the genome following BrdU addition at different time points. Orange circles along the x-axis mark sites designated as*

origins of replication by PU-seq (Daigaku et al. 2015)*. Vertical grey bars mark the centromeres. Windows were filtered to remove signal spikes by excluding those bins with a Z score > 2.5 and each dataset was normalised to account for differences in pooling (Table 3.2). Samples labelled as "Harvested 90 min after release" have the time at which BrdU had been added to the culture relative to the release from the G1 arrest (e.g. BrdU at -20 min had BrdU added 20 min prior to the release). All three chromosomes were plotted, with a zoomed in plot of 500 kb on chromosome III to aid visualisation.*

broad peaks indicative of particular initiation sites. Thus, addition of a lower concentration of BrdU to the culture did not result in localised enrichment proximal to annotated replication origin sites. Furthermore, addition of BrdU upon or following release into S phase did not result in notable incorporation during replication. This suggested the inability of cells to either import and/or process the BrdU for incorporation under conditions of ongoing replication.

3.3.3 Addition of 0.5 µM BrdU during a G1 arrest, followed by an HU arrest

To attempt to limit BrdU incorporation to only the earliest replicating regions, hydroxyurea (HU) was added concomitantly with BrdU. HU can be used to effectively arrest cells in early S phase by significantly slowing replication forks through its inhibition of ribonucleotide reductase, an enzyme required for the production of dNTPs (Singh and Xu 2016). Thus, *cdc10-m17, adh:hENT1, adh:hsv-TK* cells were arrested in G1 phase by incubation at the restrictive temperature, as in previous experiments. After 3 h 15 min at the restrictive temperature, 0.5 µM BrdU and 20 mM HU were added. This HU concentration was selected to maximally impede fork progression. Although it was higher than that typically used for cell cycle synchronisation in *S. pombe* (Luche and Forsburg 2009), the early S phase block was predicted to be reversible due to the relatively short exposure time. Following a further 20 min at the restrictive temperature, cells were released from the G1 arrest by cooling to the permissive temperature. This allowed cells to enter S phase, but the presence of HU in the medium prevented further progression (Figure 3.8). To promote thymidine and limit BrdU incorporation by replication forks upon release from HU, a tenfold excess of thymidine was

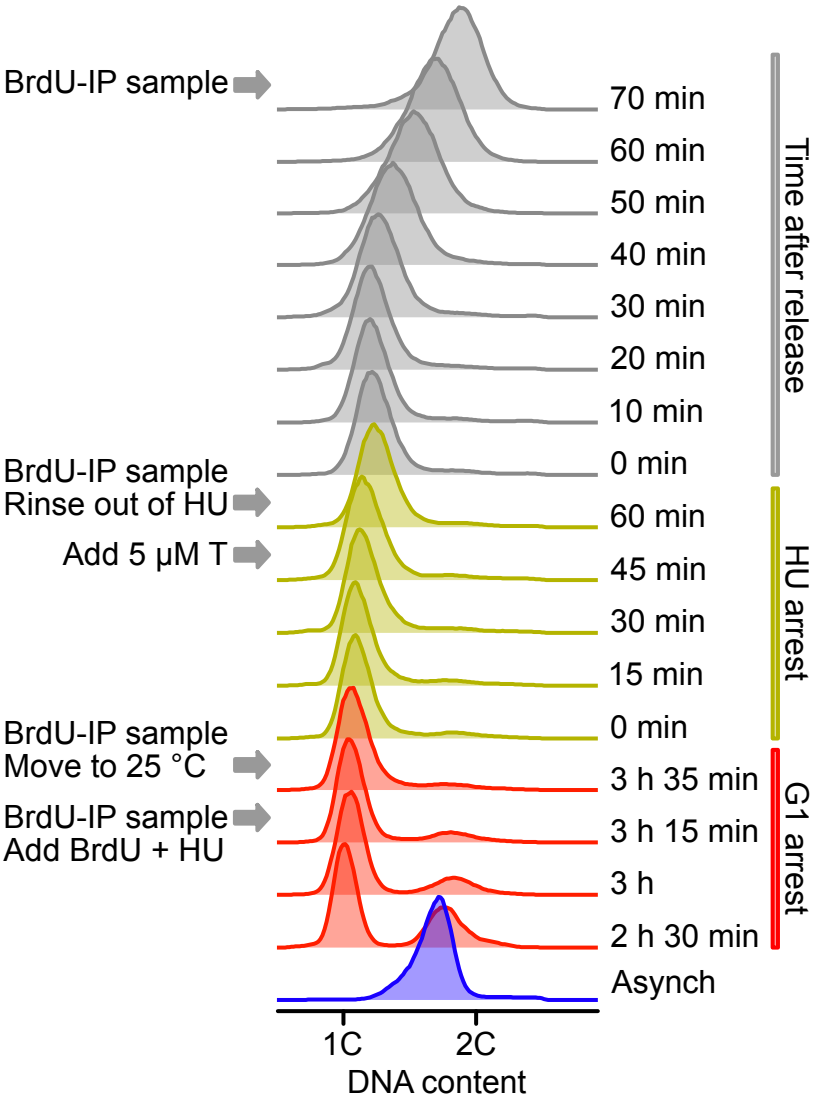


***Figure 3.8**: Histograms of DNA content during a G1 arrest, HU arrest, and release experiment in a* cdc10-m17, adh:hENT1, adh:hsv-TK *strain with 0.5 µM BrdU addition. Cells were arrested with a DNA content of 1C by incubation at the restrictive temperature. After 3 h 15 min at the restrictive temperature, a control sample was harvested for BrdU-IP-seq and 0.5 µM BrdU and 20 mM HU were added to the remaining culture. After a further 20 min at the restrictive temperature, another control sample was harvested for BrdU-IP-seq and the remaining cells were shifted to the permissive temperature, still in the presence of HU. After 45 min in HU at the permissive temperature, 5 µM thymidine was added to promote its incorporation by replication forks upon release from HU. After a total of 60 min in HU at the permissive temperature, a sample was harvested for BrdU-IP-seq before the remaining culture was rinsed into fresh medium supplemented with 5 µM thymidine and incubated at the permissive temperature to allow progression through S phase. The final sample was harvested 70 min*

after the release from HU when the majority of the population had reached a DNA content of 2C.

added to the culture 45 min after release from G1. Following a further 15 min in HU, with BrdU and thymidine, cells were rinsed and resuspended in fresh medium supplemented with an excess of thymidine. The population showed synchronous progression through S phase following the release from HU. The final sample was harvested 70 min after the release from HU as this had previously been shown to be a sufficient length of time for the population to complete genome replication (Kim & Huberman, 2001). While the majority of cells appeared to have reached a DNA content of approximately 2C in this experiment, it was decided that a release time of 90 min would be used in future HU experiments to try and maximise the proportion of the population that had completed genome replication.

DNA was extracted from the cells harvested throughout the experiment, barcoded, pooled, and subjected to BrdU-IP and Illumina sequencing. As before, the contribution of each barcode to the IP pool was normalised to the corresponding input sample (Table 3.3). This showed that approximately 45% of the IP reads were derived from the sample harvested 60 min after release from G1, when the cells were still in HU, while nearly 51% came from the final time point, 130 min after release from G1. The small increase in BrdU-IP reads between these two time points suggested that only limited BrdU incorporation occurred following release from HU, likely due to intracellular BrdU pools that could not be removed during the rinse.

To determine whether the incorporation of BrdU was indeed limited to only the earliest replicating regions by the presence of HU, the read count per 5 kb window was assessed for all three chromosomes (Figure 3.9). The profiles showed less noise than those from experiments performed without HU (Figures 3.5 and 3.7) and distinct signal peaks were observed when looking in a smaller window of the genome (lower plot). Importantly, the two samples harvested after release from the G1 arrest (60 min and 130 min) showed similar

***Table 3.3**: Summary of read counts for Illumina libraries. Barcoded input and IP samples were pooled and a unique index was added during amplification of each pool. The barcodes added to each sample prior to pooling were used to assess the number of reads for each sample within the pools. The 3 h 15 min sample was the same as that shown for the 1 μM BrdU addition experiment (section 3.3.1) as this experiment was performed in parallel, from the same starting culture. The starting culture was divided between the two experiments before the addition of BrdU (and HU in this case). The fraction of each sample in the IP pool was normalised to the corresponding input to account for differences in the sample pooling prior to IP.*

Index pool	Sample	Read count	Fraction of reads in pool	Input-normalised fraction of reads in pool
Input	3 h 15 min in G1 arrest	183783	0.310	NA
	3 h 35 min in G1 arrest	163852	0.276	NA
	60 min after G1 release, in HU	129757	0.219	NA
	130 min after G1 release (70 min after HU release)	115785	0.195	NA
IP	3 h 15 min in G1 arrest	3131	0.016	0.011
	3 h 35 min in G1 arrest	8407	0.042	0.032
	60 min after G1 release, in HU	93194	0.469	0.449
	130 min after G1 release (70 min after HU release)	93947	0.473	0.508

patterns of BrdU incorporation, which were different to those of the pre-BrdU and pre-release samples.

To test whether the peaks in these data corresponded to known initiation sites that were identified by polymerase-usage sequencing (PU-seq) (Daigaku et al. 2015), peaks were called as 1 kb windows in which the number of IP reads was greater than 75, following Z score filtering and input normalisation. These peaks were then compared with PU-seq initiation sites and the

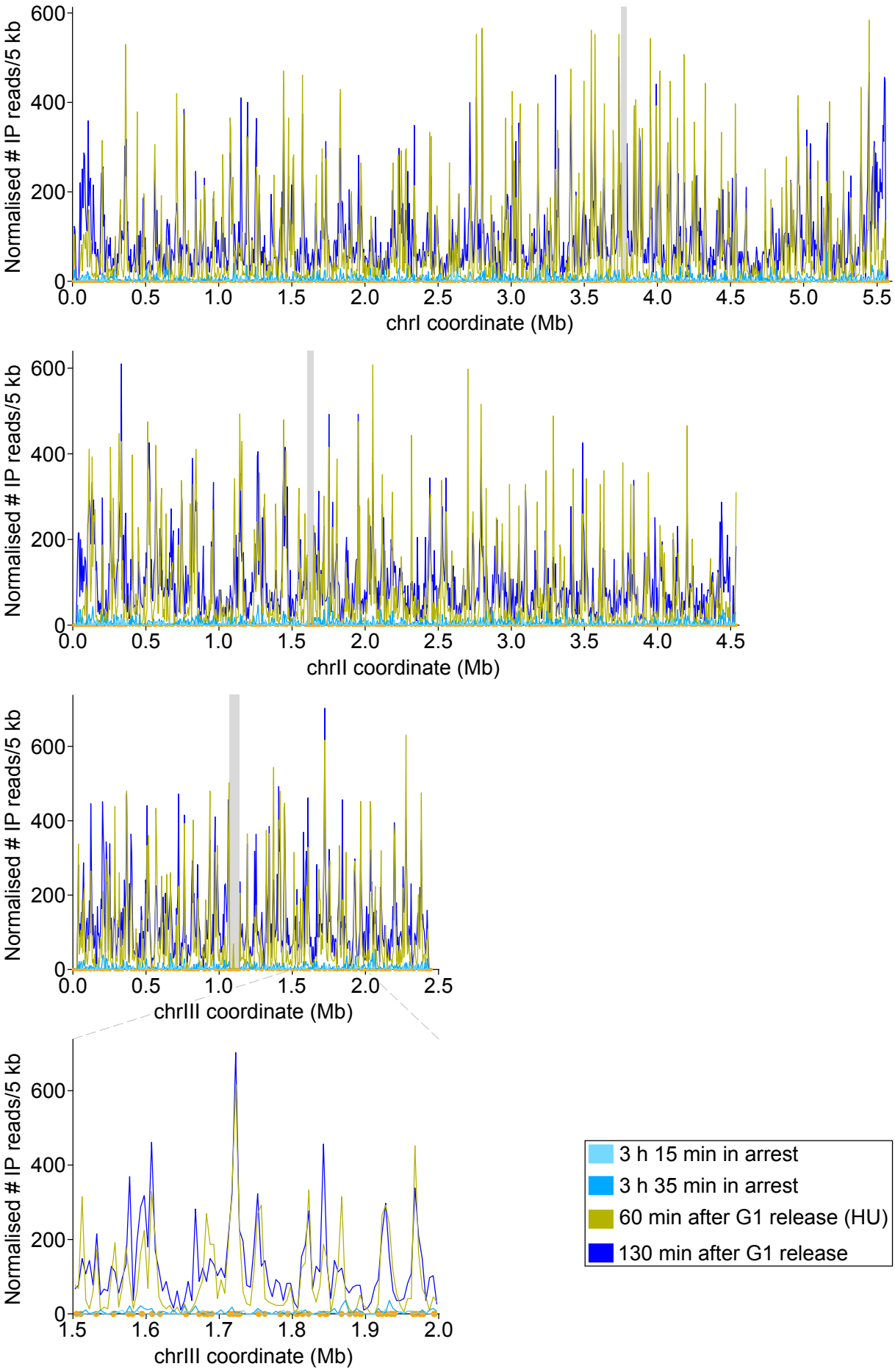
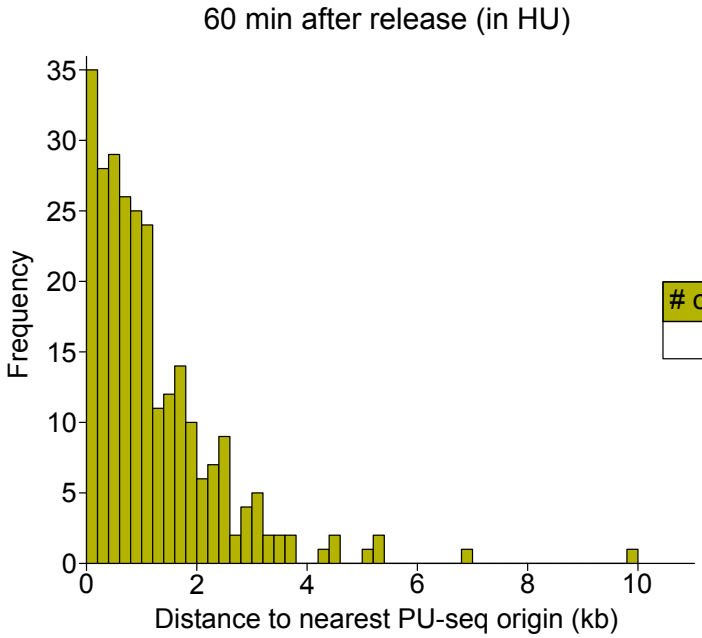


***Figure 3.9**: Number of BrdU-IP reads mapping to 5 kb bins across the genome. Orange circles*

along the x-axis mark sites designated as origins of replication by PU-seq (Daigaku et al. 2015)*. Vertical grey bars mark the centromeres. Windows were filtered to remove signal spikes by excluding those bins with a Z score > 2.5 and each dataset was normalised to account for differences in pooling (Table 3.3). All three chromosomes were plotted, with a zoomed in plot of 500 kb on chromosome III to aid visualisation.*

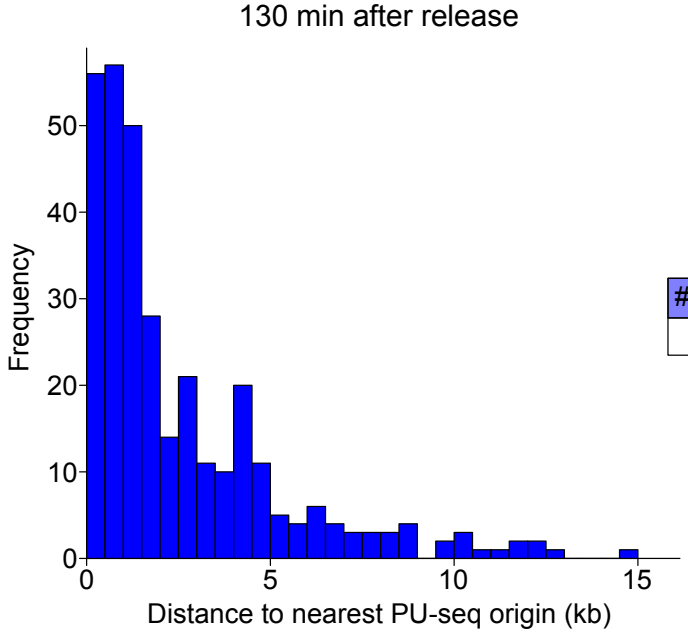
distribution of the distances between each BrdU peak and its nearest PU-seq initiation site was plotted (Figure 3.10). This was performed for both the data from the sample harvested 60 min after the release from G1, immediately before release from HU (Figure 3.10 A), and the data from the sample harvested 130 min after the G1 release (70 min after release from HU, Figure 3.10 B). In addition, as a control, an intermediary number of random 1 kb bins were likewise compared to PU-seq origins (Figure 3.10 C). When analysing the data from the sample harvested 60 min after the release from G1, 261 peaks were identified, with a median distance of 850 bp to the nearest PU-seq initiation site. The same analysis on the data from the sample harvested 130 min after the G1 release found 323 peaks of BrdU signal, with a median distance of 1,450 bp to the nearest PU-seq initiation site. Thus more peaks were detected at this later time point, and a higher fraction of the detected peaks were further from annotated initiation sites. This was consistent with some BrdU incorporation by elongating forks after release from HU. Nevertheless, the high proportion of BrdU peaks detected within 2 kb of a PU-seq initiation site indicated that early initiation events were preferentially labelled with BrdU in this experimental setup. Analysis of 291 randomly generated 1 kb windows, yielded a median distance of 3,268 bp to the nearest PU-seq origins, greater than both medians for the experimental data. This was consistent with a genuine enrichment proximal to annotated initiation sites in the BrdU-IP data. Therefore, it was concluded that this approach combined with D-NAscent might allow detection of initiation events on single molecules.

A



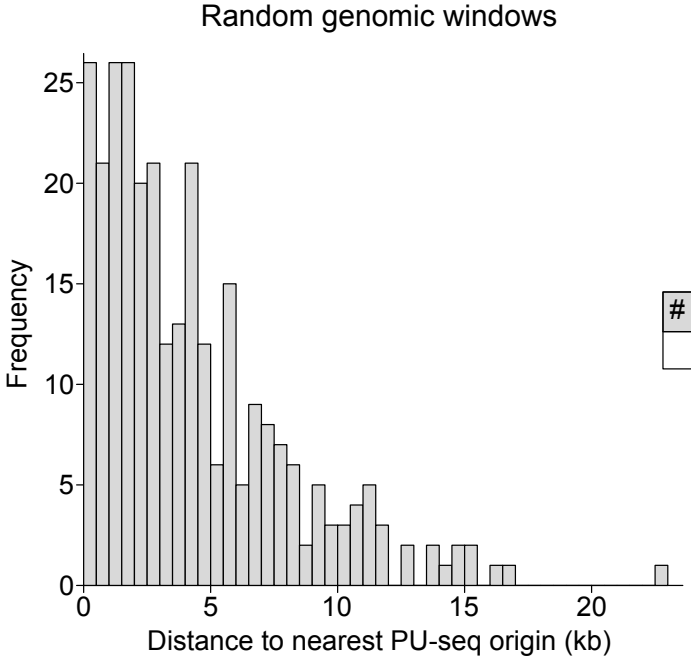
# of peaks	Mean (bp)	Median (bp)
261	1,224	850

B



# of peaks	Mean (bp)	Median (bp)
323	2,584	1,450

C



# of peaks	Mean (bp)	Median (bp)
291	4,272	3,268

***Figure 3.10**: Histograms of the distances between peaks in BrdU-IP signal and PU-seq initiation sites. BrdU peaks were defined as 1 kb windows that had more than 75 IP reads. (**A**) Peaks detected at 60 min after release from G1, when cells were still arrested in HU were compared with PU-seq initiation sites. A total of 261 BrdU-IP peaks were found through this analysis, with a median distance of 850 bp to the nearest PU-seq initiation site. (**B**) Peaks detected at 130 min after release from G1 (70 min after release from HU) were compared with PU-seq initiation sites. The number of peaks detected increased to 323, as did the median distance to the nearest initiation site (1,450 bp). (**C**) The same analysis of 291 random 1 kb genomic windows showed a slightly different distribution to the experimental data, with a greater median distance of 3,268 bp to the nearest PU-seq origin. Note the different x- and y-axis scales in the different panels.*

In conclusion, BrdU-IP-seq experiments confirmed BrdU incorporation during replication after release from a G1 arrest. These population datasets showed that the IP signal was uniformly distributed across the genome following a G1 arrest, consistent with observations in a G2 arrest/release strain assessed by deep sequencing (Xu et al. 2012). Addition of HU during G1 did result in more localised BrdU incorporation proximal to previously-identified initiation sites. Thus, this experimental design with HU addition was used for subsequent experiments in which DNA would be analysed by D-NAscent.

3.4 Validation of molecule length by PFGE

Having tested protocols for cell cycle synchronisation and BrdU incorporation in the *cdc10-m17, adh:hENT1, adh:hsv-TK* strain, the next requirement was to assess the molecule size distribution following careful DNA extraction from *S. pombe* cells. To this end, pulsed-field gel electrophoresis (PFGE) was used to compare the molecule sizes of DNA extracted from a wild-type *S. pombe* strain to known *S. pombe* and *S. cerevisiae* chromosomal markers. The extraction method used was not anticipated to yield *S. pombe* chromosomal-length molecules, but the aim was to assess whether the extraction gave molecules of sufficient length for effective analysis of replication dynamics by nanopore sequencing. The PFGE run conditions

had to be optimised to provide maximal separation and thus accurate sizing of the extracted molecules. Initially, Megabase agarose was used with a run time of 24 h, as recommended for separation of *S. cerevisiae* chromosomes in the PFGE mapper manual. Upon completion of the run, ethidium bromide staining revealed that the *S. cerevisiae* chromosomal markers had

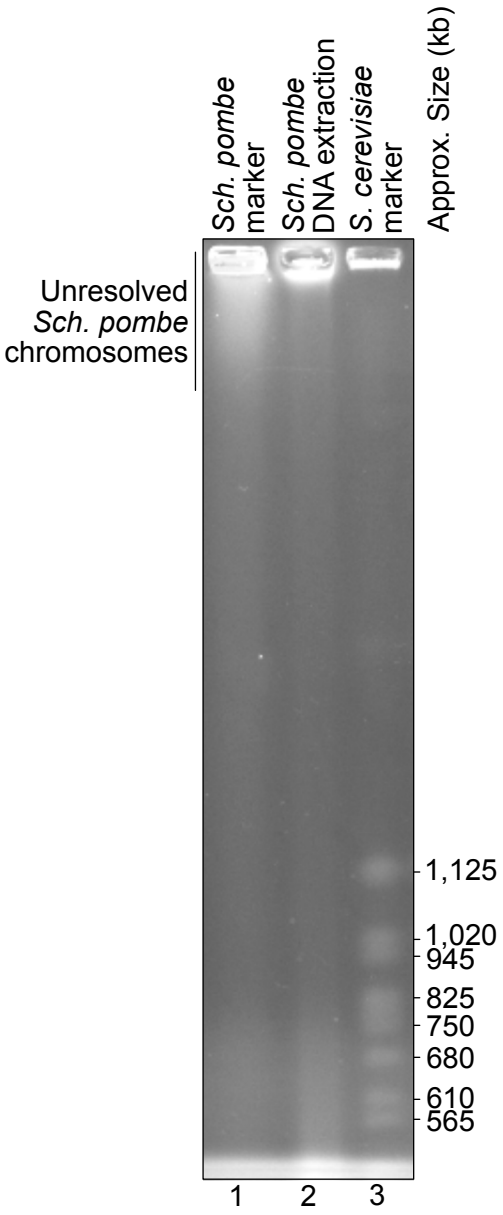


Figure 3.11*: PFGE trial to find suitable run conditions to assess* S. pombe *DNA extraction. Genomic DNA was run on a Megabase agarose gel for 24 h, along with both* S. pombe *and* S. cerevisiae *chromosomal markers. The conditions used were unsuitable for resolving the* S. pombe *chromosomes, while the* S. cerevisiae *chromosomes shorter than 565 kb ran off the end of the gel.*

partially run out of the end of the gel (Figure 3.11). The *S. pombe* chromosomal markers did not show any separation, indicating that these run conditions were inappropriate for size separation of *S. pombe* chromosome-length molecules. Signal corresponding to the extracted *S. pombe* genomic DNA was localised in the lowest part of the gel. Together, these observations showed that the run conditions used were not suitable to visualise molecules < 565 kb.

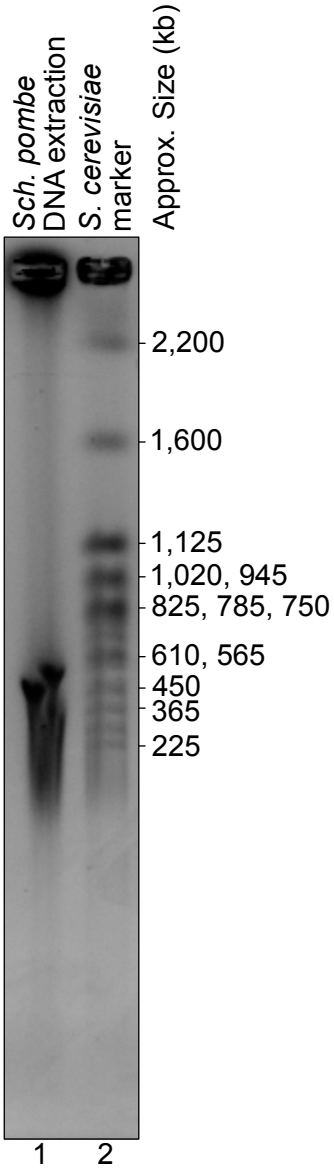


***Figure 3.12**: Second PFGE trial to find suitable run conditions to assess* S. pombe *DNA extraction. A gel was run as in Figure 3.11 but for half the run time (12 h).* S. cerevisiae *chromosomal markers were sufficiently resolved to estimate the* S. pombe *genomic DNA to* < *600 kb in length.*

To enable size separation and subsequent visualisation of molecules in the range of the *S. cerevisiae* chromosomal markers, a PFGE experiment was performed with a run time of 12 h, rather than the 24 h used previously. Staining the total DNA showed that all *S. cerevisiae* chromosomal markers were retained within the gel, however, the size separation was sub-optimal (Figure 3.12). This indicated that a longer run time (somewhere between 12-24 h)

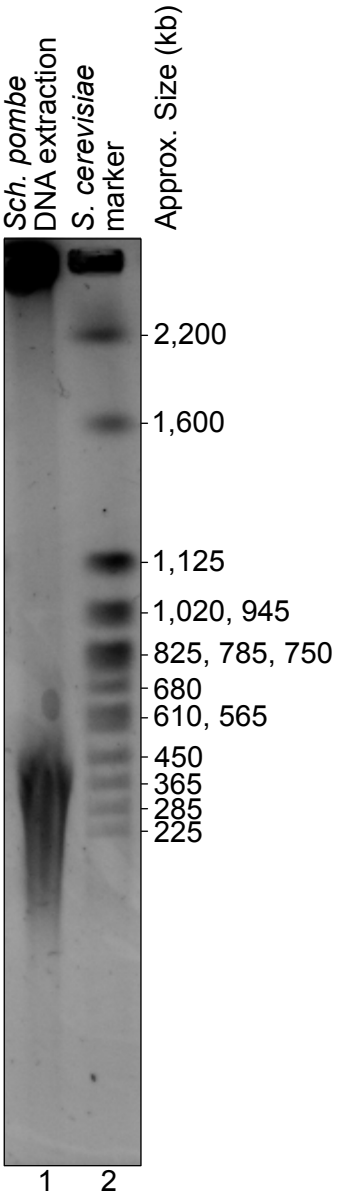


Figure 3.13*: Third PFGE trial to find suitable run conditions to assess* S. pombe *DNA extraction. A gel was run as in Figures 3.11 and 3.12 with two changes. First, the gel was made using SeaKem LE agarose, rather than Megabase agarose. Second, the total run time was 16 h in this trial. The* S. cerevisiae *chromosomal marker was more clearly resolved and the majority of the* S. pombe *genomic DNA appeared to be < 450 kb in length.*

would allow adequate separation of molecules within this size range. In addition, this gel showed that the molecules extracted from *S. pombe* were mostly < 610 kb and ranged downwards in size to < 225 kb, the smallest *S. cerevisiae* chromosomal marker.

The PFGE was repeated to confirm the previous results and with the aim of achieving more optimal size separation. To that end, standard agarose was used (rather than Megabase agarose) with a run time of 16 h; otherwise run conditions remained unchanged (Figure 3.13). These conditions improved the resolution of the *S. cerevisiae* chromosomal markers and confirmed that the *S. pombe* DNA sample contained molecules ranging from < 225 kb to approximately 450 kb. This indicated that careful extraction and purification yielded molecules of sufficient length for nanopore sequencing analysis.

In summary, PFGE analysis of DNA extracted from *S. pombe* cells showed that high molecular weight DNA, suitable for nanopore sequencing, could be obtained.

3.5 Assessment of BrdU incorporation by D-NAscent

After establishing experimental conditions to obtain BrdU-labelled, high molecular weight genomic DNA, the next step was to test whether these conditions were indeed appropriate for D-NAscent.

3.5.1 Addition of 0.5 µM BrdU during a G1 arrest, followed by an HU arrest

Given the promising BrdU-IP sequencing results in section 3.3.3, a biological repeat of this experiment was performed to harvest a sample for D-NAscent analysis. The final time point at which cells were harvested for DNA extraction was changed from 70 min to 90 min after the release from HU, to allow more cells to complete genome replication. BrdU was added to the culture to a final concentration of 0.5 µM, concomitantly with 20 mM HU. Flow cytometry analysis of total DNA content in cells harvested throughout the experiment indicated a successful G1 arrest, with the majority of cells showing a DNA content of 1C before release at the permissive temperature (Figure 3.14). Release into HU caused cells to accumulate in early

S phase and, after 60 min under these conditions, HU was removed and cells progressed synchronously through S phase. A tenfold excess of thymidine was added both 15 min before and at the time of release from HU to promote thymidine rather than BrdU incorporation during replication in the released cells.

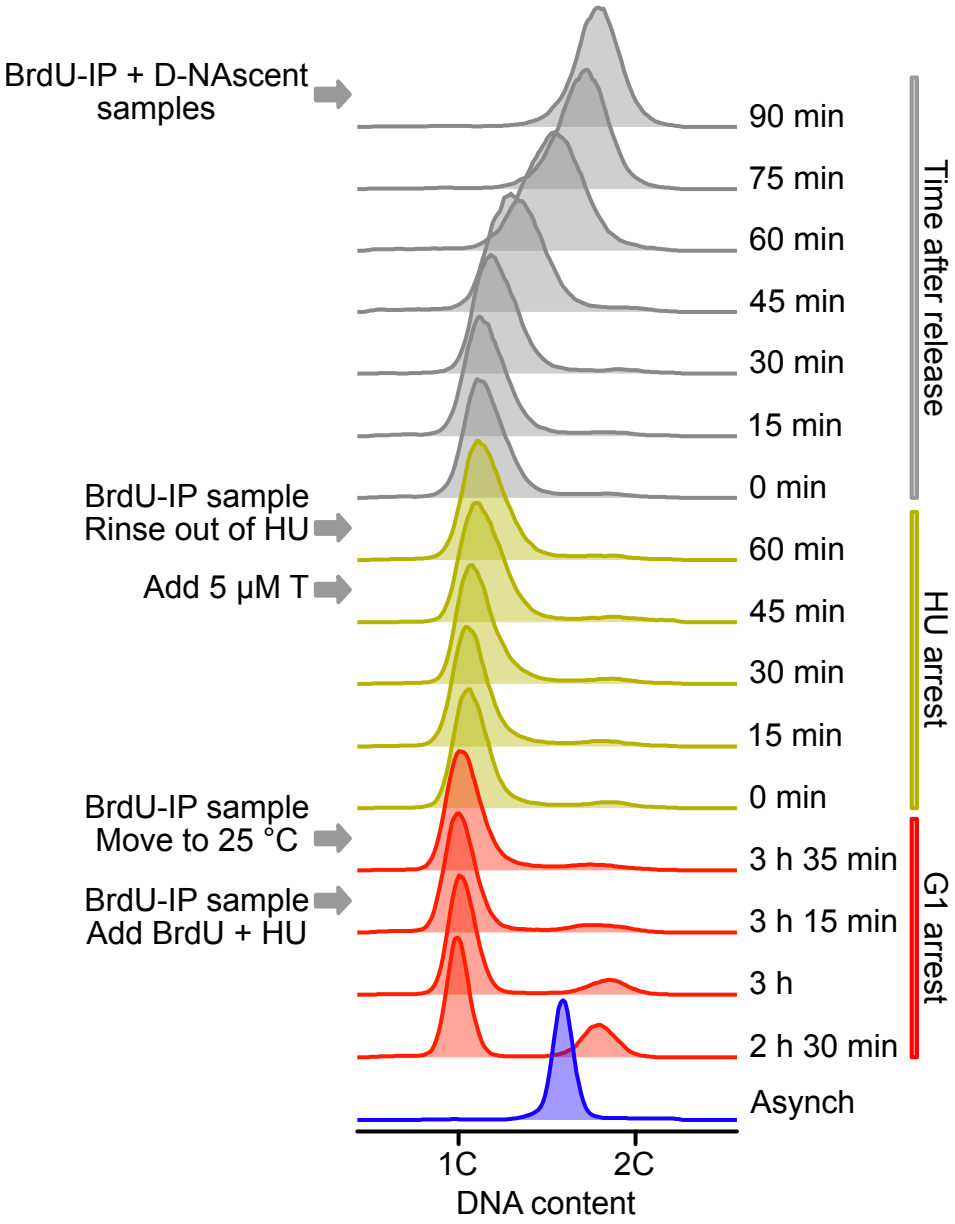


***Figure 3.14**: Histograms of DNA content during a G1 arrest, HU arrest, and release experiment in a* cdc10-m17, adh:hENT1, adh:hsv-TK *strain with 0.5 µM BrdU addition. As in Figure 3.8, with the final sample harvested for BrdU-IP and D-NAscent analysis 90 min after the release from HU (150 min after release from G1).*

The nanopore data were analysed using an updated version of the D-NAscent pipeline (DNAscent_v2) (Boemo 2020) (https://github.com/MBoemo/DNAscent). This software uses a residual neural network to assign the probability of a BrdU at each thymidine in a read, based on the nanopore signal. Thus, every thymidine in a molecule was scored between 0 and 1, where values near 0 indicated the presence of thymidine, while values close to 1 reflected a high probability that a base was BrdU. The BrdU-IP-seq samples served as a control for the D-NAscent data. To facilitate comparison between the datasets, an ensemble of the DNAscent detect output was produced and plotted across the genome with the IP data (Figure 3.15). This revealed that peaks in the two datasets predominantly coincided, and that they colocalised with PU-seq-identified initiation sites (Daigaku et al. 2015) (Figure 3.15). This provided validation of the D-NAscent approach for mapping replication dynamics in this experimental system.

Given that the goal of this work was to study replication dynamics in single molecules, the next step was to visualise the pattern of BrdU incorporation in individual nanopore reads. The reads were visualised in a genome browser, with the orange bars showing the probability of a BrdU at each thymidine along a molecule, based on the output from the DNAscent_v2 software (Figure 3.16). For comparison, a molecule with very low probability of BrdU throughout its length, which was therefore assumed to be the parental strand during replication, is also shown (Mol 1). The BrdU incorporation in the other molecules was generally lower than the optimal level for accurate prediction of single molecule replication dynamics. Therefore, it was concluded that the experimental setup was suboptimal for D-NAscent and that it was necessary to increase the levels of BrdU incorporation at early firing forks.

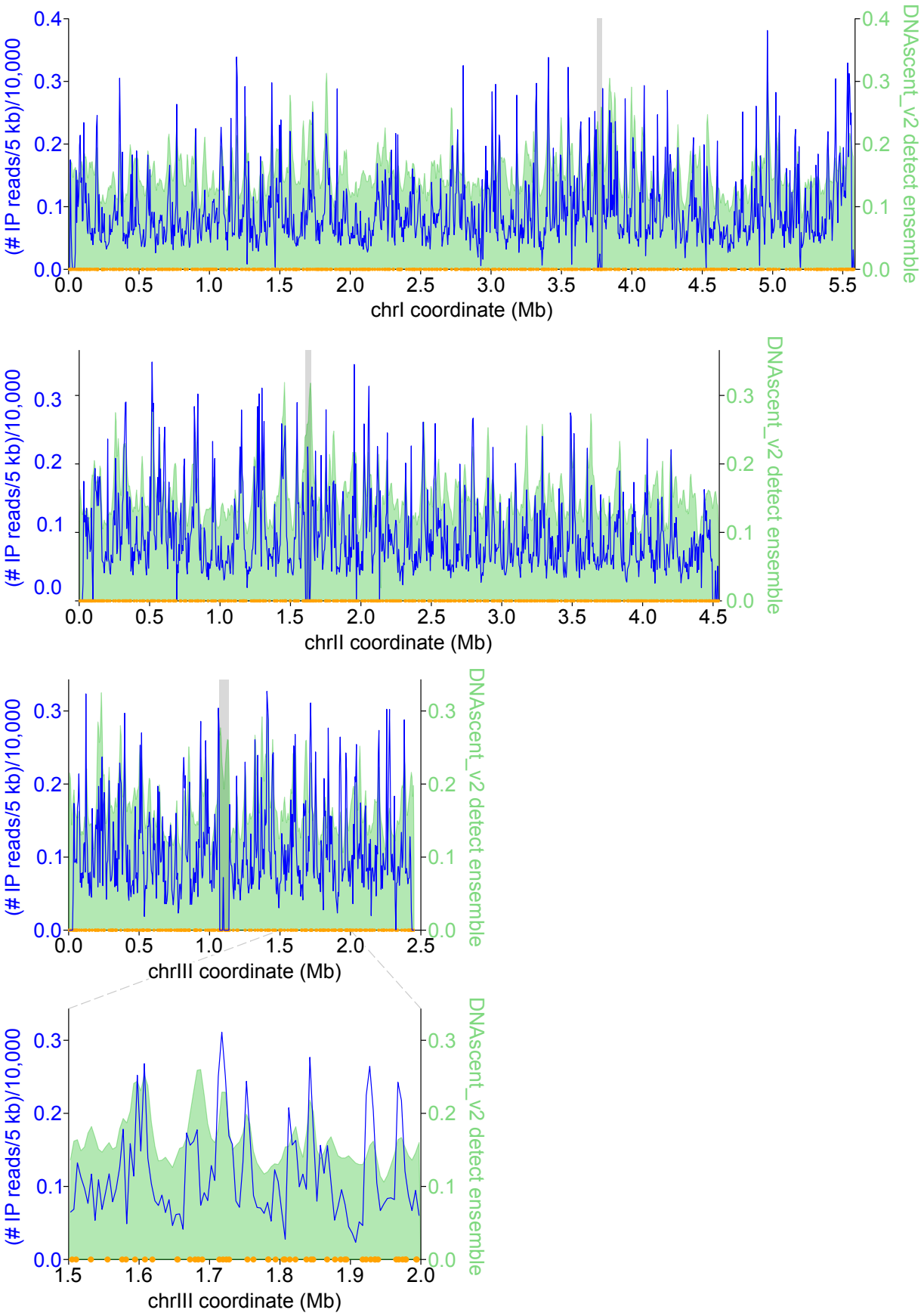


***Figure 3.15**: Visualisation of an ensemble of D-NAscent data (green) with BrdU-IP data from the same time point (blue). The D-NAscent ensemble data show the fraction of thymidines in each window that are being called as BrdU, based on a probability threshold of 0.7. The*

number of IP reads per 5 kb window was divided by 10,000 to enable plotting on a comparable scale to the D-NAscent ensemble data. Orange circles along the x axis mark replication initiation sites identified by PU-seq (Daigaku et al. 2015)*. Vertical grey bars mark the centromeres.*

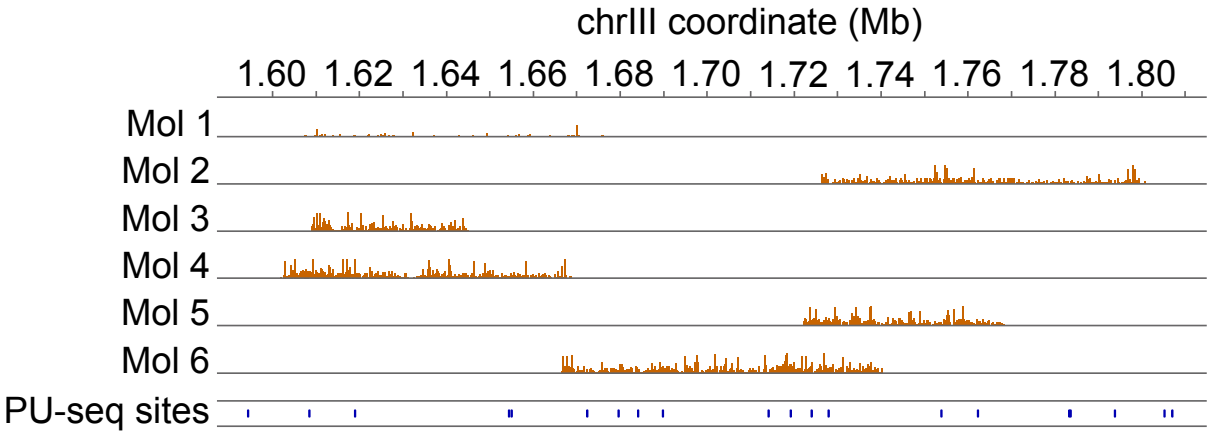


Figure 3.16*: Visualisation of example individual D-NAscent reads aligning to chromosome III. Each molecule is represented by a track showing the probability of a given thymidine being a BrdU in orange. The y-axis scale for each track ranges from 0 to 1. Known initiation sites identified by PU-seq* (Daigaku et al. 2015) *are annotated in blue.*

3.5.2 Addition of 1 μM BrdU during a G1 arrest, followed by an HU arrest

Given that the D-NAscent reads in Figure 3.16 did not show clear peaks of incorporated BrdU, as anticipated for early initiation events, flanked by regions of lower BrdU levels as forks incorporated more thymidine, it was necessary to try to optimise BrdU incorporation levels in nascent strands. Thus, the experiment described in section 3.5.1 was repeated, but with addition of 1 μM BrdU (Figure 3.17). Based on the total DNA content, cells showed similar arrest and release behaviour as seen previously when exposed to HU (Figures 3.8 and 3.14).

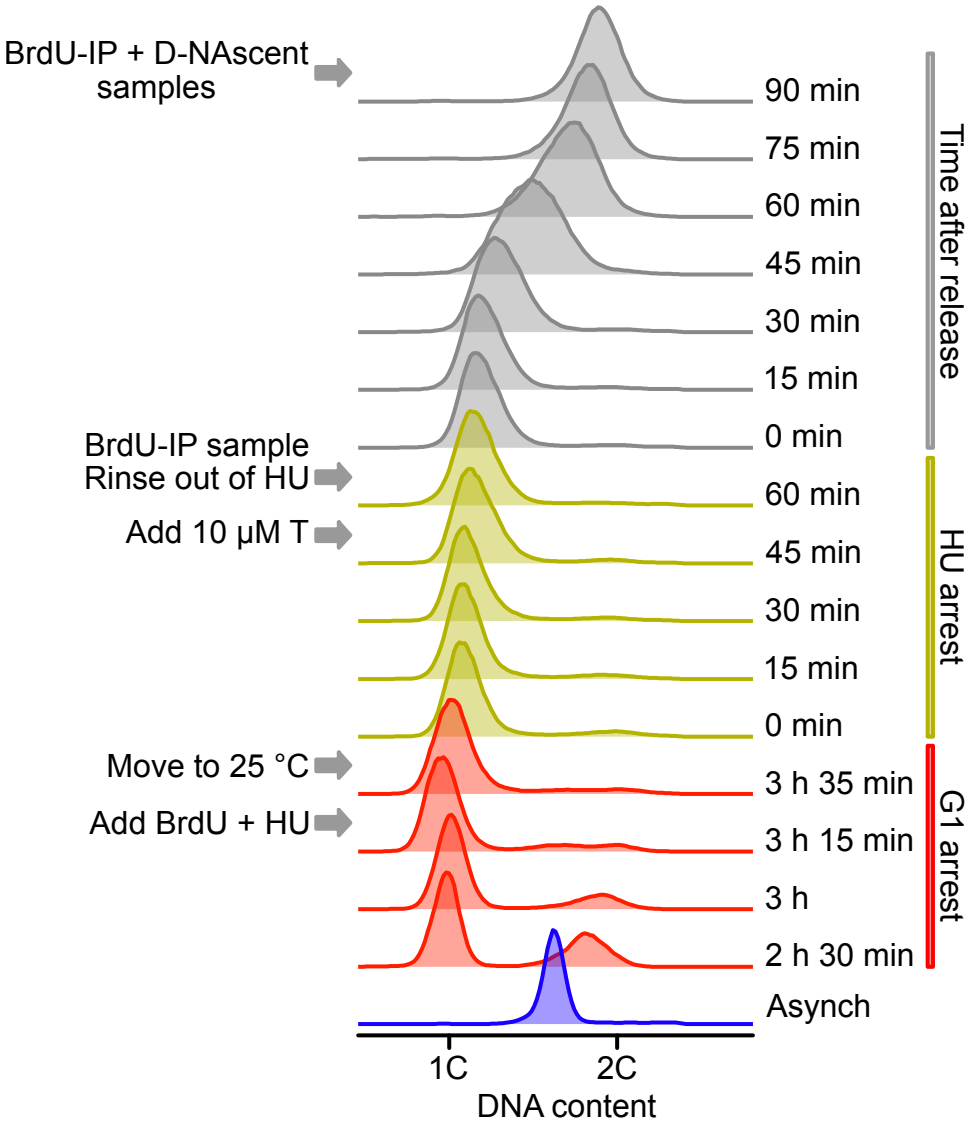


***Figure 3.17**: Histograms of DNA content during a G1 arrest, HU arrest, and release experiment in a* cdc10-m17, adh:hENT1, adh:hsv-TK *strain, as in Figure 3.14, with 1 μM BrdU addition.*

As before, BrdU-IP and D-NAscent analyses were performed on the DNA extracted from cells harvested 90 min after release from the HU arrest. To compare the two approaches for analysing replication dynamics, an ensemble of the single-molecule data was compared to the BrdU-IP data for this sample (Figure 3.18). The general agreement in the peak locations between the two datasets supported the validity of the D-NAscent approach, as the ensemble of its output reflected that of the established BrdU-IP methodology. Furthermore, the addition of 1 μM BrdU seemed to result in clearer peaks in the D-NAscent ensemble profile, compared to addition of 0.5 μM (Figure 3.15). In addition, the overall signal was greater than in Figure 3.15, consistent with the doubling of the BrdU concentration.

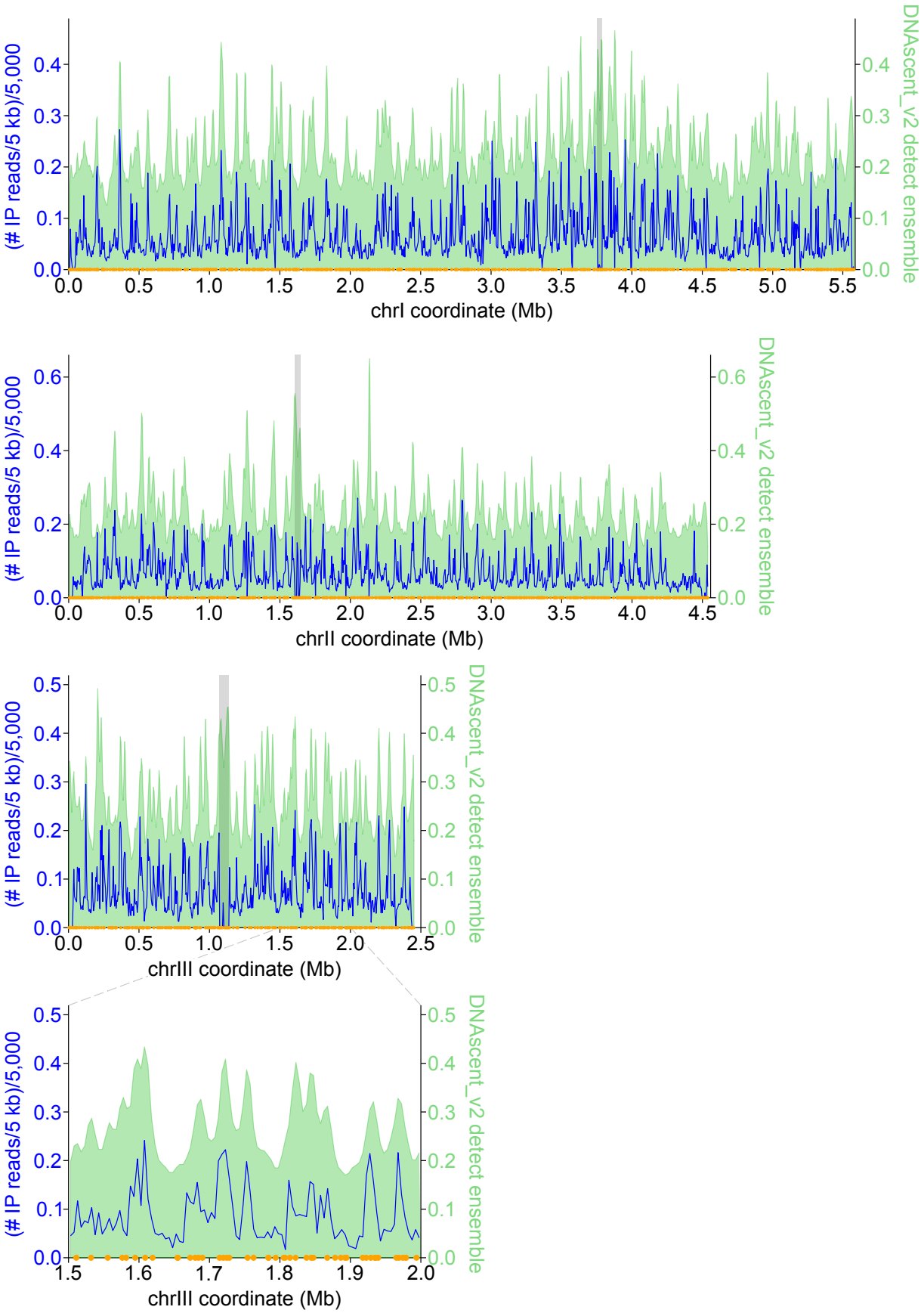


***Figure 3.18**: Visualisation of an ensemble of D-NAscent data (green) with BrdU-IP data from the same time point (blue), as in Figure 3.15. The number of IP reads per 5 kb window was divided by 5,000 to enable plotting on a comparable scale to the D-NAscent ensemble data.*

Orange circles along the x axis mark replication initiation sites identified by PU-seq (Daigaku et al. 2015)*. Vertical grey bars mark the centromeres.*

To assess whether the BrdU signal was able to reveal the replication dynamics of single molecules in these data, sequencing reads analysed by D-NAscent were visualised in a genome browser (Figure 3.19), as above. For the interpretation of replication dynamics from such data, the anticipated pattern of BrdU incorporation would be localised peaks of high BrdU probability detected by the software, flanked on both sides by a decrease in BrdU signal to background levels. This would be consistent with BrdU incorporation at replication forks that fired and elongated for a limited stretch in the presence of HU before reversibly stalling until release from the HU arrest in the presence of a thymidine chase. An additional functionality in DNAscent_v2 is forkSense, which predicts the probabilities of leftward- and rightward-moving forks across the molecules, based on the BrdU signal. Thus, for each single molecule, three tracks of information were displayed – the probability of BrdU at each thymidine position (track a), the probability of a leftward-moving fork (track b) and the probability of a rightward-moving fork (track c). Again, Mol 1 showed the anticipated BrdU probabilities for a parental strand. There was an indication of peaks in BrdU signal in Mol 2 and Mol 4 (Figure 3.19) in the region between chromosomal coordinates 1.70 – 1.75 Mb, suggesting that initiation events may have taken place at these sites. The forkSense analysis did detect these potential BrdU peaks, and displayed an inflection between the predictions for leftward- and rightward-moving forks at these locations. Such inflections are interpreted by the forkSense software as initiation sites. It is clear from these example molecules that, while the potential peaks in Mols 2 and 4 are likely to be accurate initiation site predictions by forkSense, the pattern of BrdU incorporation in the other molecules, and even elsewhere within these two molecules was not conducive to reliable predictions of replication dynamics. Therefore, the other inflection points in the forkSense tracks seem unlikely to represent genuine initiation events in these data.

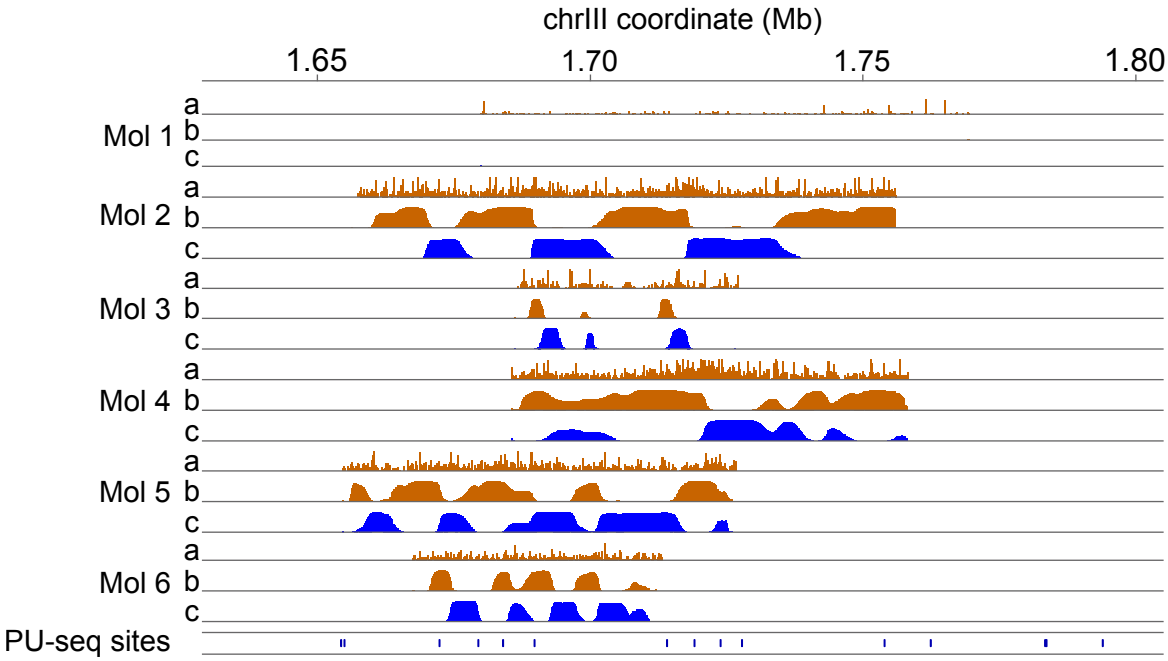


***Figure 3.19**: Visualisation of individual D-NAscent reads aligning to chromosome III. Each molecule is represented by three tracks: a shows the probability of a given thymidine being BrdU, b shows the predicted probability of a leftward-moving fork replicating a region, while c shows the predicted probability of synthesis by a rightward-moving fork. The y-axis scale for each track ranges from 0 to 1. Known initiation sites identified by PU-seq* (Daigaku et al. 2015) *are annotated.*

To test whether the forkSense initiation site predictions matched the expectations based on previously published replication timing data, a pileup of all forkSense-determined initiation events was plotted across the genome, along with the corresponding BrdU-IP data, as well as PU-seq origins and replication timing data from both sort-seq and cell cycle synchronisation by elutriation (Daigaku et al. 2015) (Figure 3.20). Interestingly, it appeared that almost any site on each chromosome acted as an origin of replication according to the forkSense analysis.

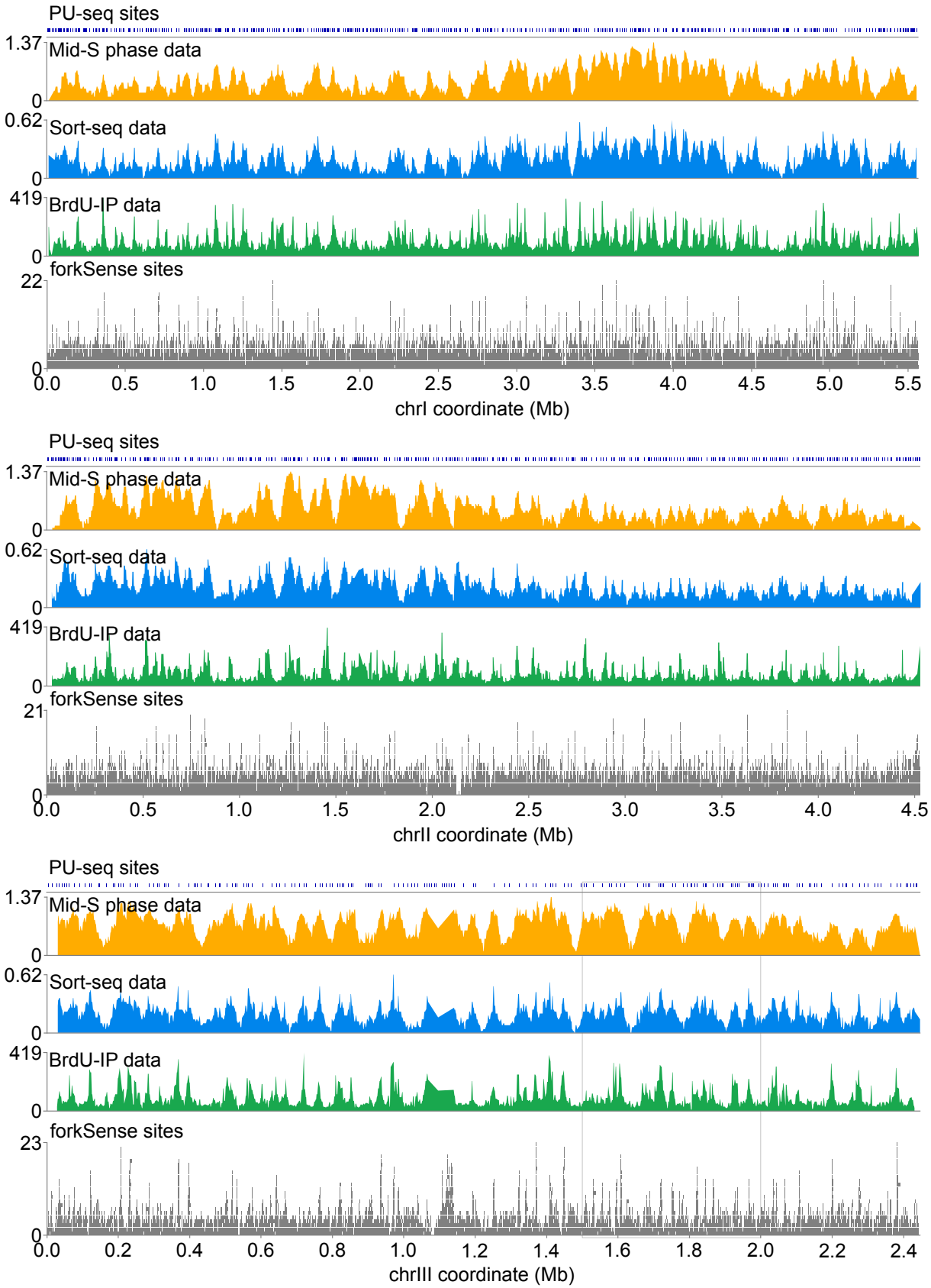


Figure 3.20*: Sites of forkSense-assigned initiation events across all three chromosomes. Each grey bar indicates an individual initiation event predicted by forkSense. The BrdU-IP data from the same sample were plotted in green, while replication timing data from either sort-seq or*

cell cycle synchronisation (Daigaku et al. 2015) *were plotted in blue and orange, respectively. The y-axis for each of these datasets is as follows: forkSense data – number of origin calls, BrdU-IP data – number of IP reads per 1 kb bin, sort-seq and mid-S phase data – relative copy number. The grey box around chromosome III coordinates 1.5-2.0 Mb highlights the region visualised in Figure 3.21.*

Closer inspection of a 500 kb window on chromosome III (Figure 3.21) indicated that the regions where the most origins were predicted by forkSense generally colocalised with peaks in replication timing and BrdU incorporation.

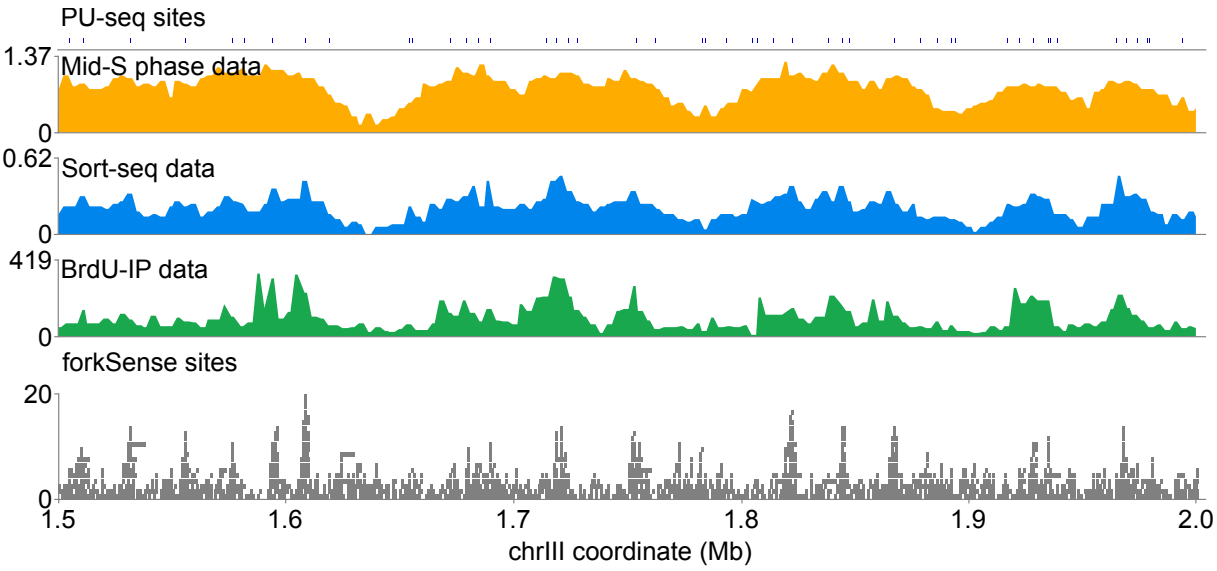


Figure 3.21*: A zoom-in on a 500 kb window of chromosome III from Figure 3.20 to aid visualisation.*

To test whether the forkSense-predicted origin locations matched those detected by PU-seq (Daigaku et al. 2015), the distance between each of the 46,390 forkSense origins (detected in 28,867 molecules) and their nearest PU-seq origin was plotted (Figure 3.22). In addition, an equal number of random genomic windows was generated for the same comparison to the PU-seq origins. As each forkSense origin varied in length, the median length of each site was calculated (539 bp) and this was used as the bin size for the random sites. The majority of the forkSense origins were within 5 kb of a PU-seq origin, with a median distance of 2,116 bp.

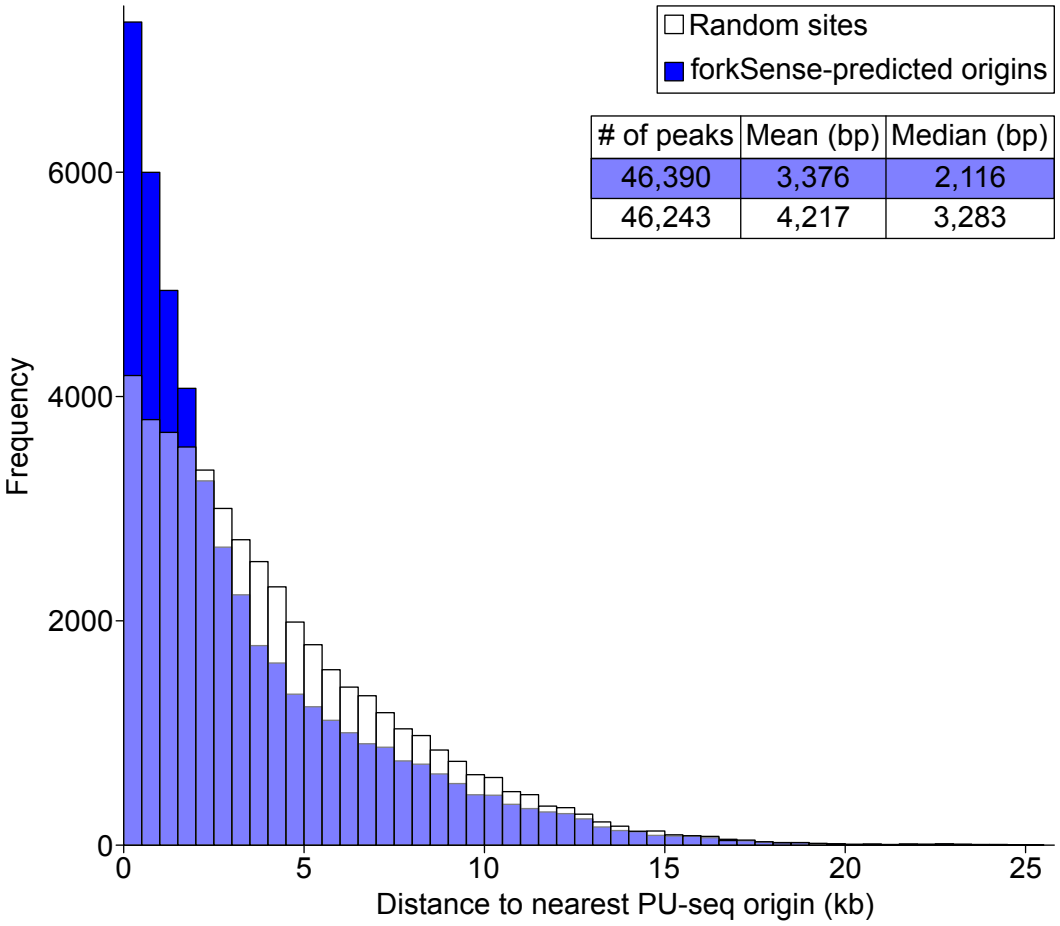


***Figure 3.22**: The distribution of distances between each forkSense-predicted origin and its nearest PU-seq origin* (Daigaku et al. 2015) *(blue). A total of 46,390 initiation events had been detected across all molecules greater than 30 kb in length by forkSense. The median distance between these and the PU-seq origins was 2,116 bp. In white, the distribution of distances between 46,243 randomly-generated 539 bp windows and their nearest PU-seq origins were plotted. The median distance between these random windows and the PU-seq origins was 3,283 bp.*

Meanwhile, the randomly-generated windows showed a slightly different distribution, with a median distance of 3,283 bp to the nearest PU-seq origin. Therefore, the forkSense origin calls were only slightly closer to PU-seq origins than the random sites.

In summary, based on visualisation of the single molecules, the addition of 1 μM BrdU resulted in higher levels of BrdU incorporation than 0.5 μM (section 3.5.1), as anticipated. Furthermore, the locations of each of the initiation sites predicted by forkSense were predominantly in fairly

close proximity to a PU-seq origin. However, the single molecule data for this experiment indicated that clearer transitions between BrdU positive and BrdU negative regions were relatively infrequent under these conditions. Therefore, while the forkSense origin calls generally matched expectations, further optimisations of the setup were required.

3.5.3 Addition of 1 μM BrdU during a G1 arrest

To investigate replication dynamics in single molecules from cells with minimal perturbation to S phase, a G1 arrest and release was performed without addition of HU. BrdU was added to a final concentration of 1 μM 3 h 15 min into the arrest. After 20 min in the presence of BrdU

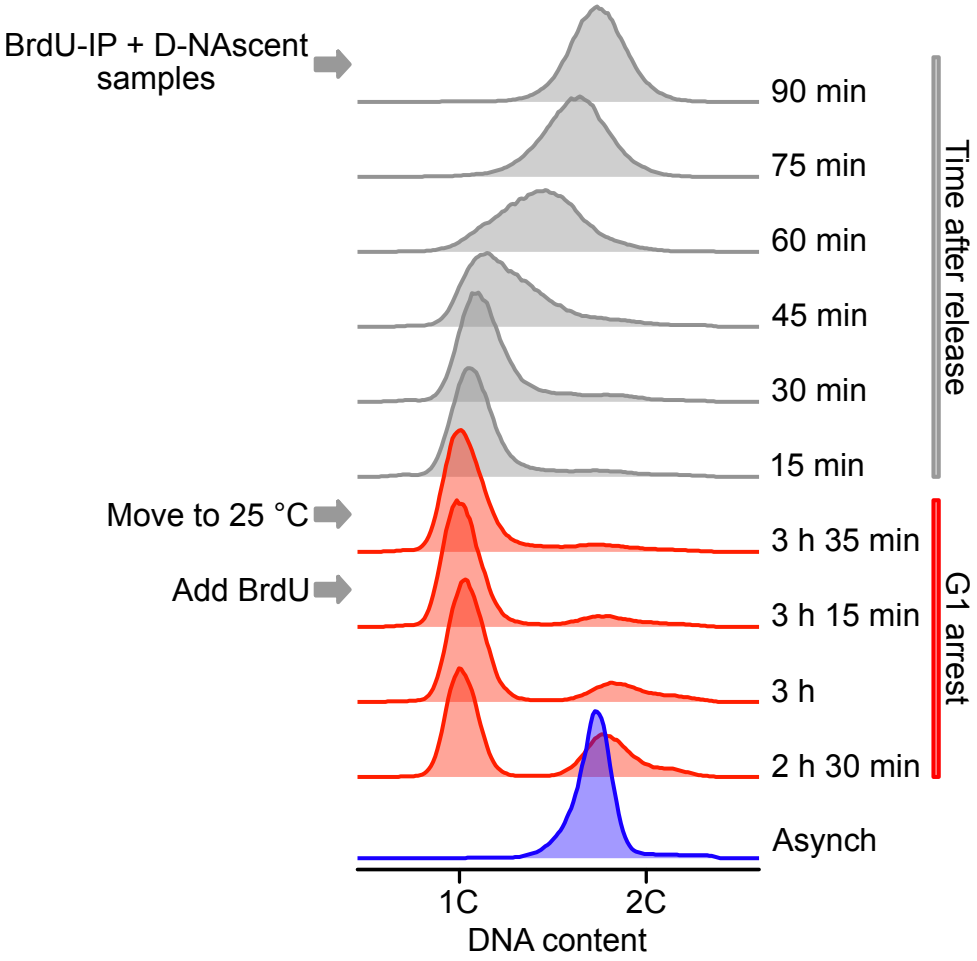


***Figure 3.23**: Histograms of DNA content during a G1 arrest and release experiment in a* cdc10-m17, adh:hENT1, adh:hsv-TK *strain with 1 μM BrdU addition during the arrest, as in Figure 3.4. Cells were harvested for D-NAscent and BrdU-IP analysis 90 min after release from the G1 arrest.*

at the restrictive temperature, cells were released from G1 by cooling to the permissive temperature and incubated for a further 90 min to allow S phase progression. Flow cytometry analysis of the bulk genomic DNA content of cells confirmed the cell cycle synchronisation (Figure 3.23).

DNA from the cells harvested 90 min after the release was analysed by both BrdU-IP-seq and D-NAscent. As in Figure 3.18, an ensemble of the D-NAscent detect data was plotted with the BrdU-IP data (Figure 3.24). Both datasets showed relatively uniform BrdU incorporation across the genome, consistent with the findings in sections 3.3.1 and 3.3.2, and in contrast to the HU experiments in sections 3.3.3, 3.5.1, and 3.5.2. The lack of clear BrdU-peaks in either dataset indicated that the incorporation of BrdU was relatively even genome-wide, as opposed to preferential at some sites (e.g. early-replicating regions). It was therefore concluded that the intracellular BrdU concentration did not change as cells progressed through S phase and thus BrdU was incorporated by replication forks regardless of replication time.

To assess the impact of removing the HU arrest on the replication dynamics in single molecules, reads were visualised as in Figures 3.16 and 3.19. The increase in BrdU probability across the molecules was consistent with greater and more wide-spread incorporation than in previous experiments (Figure 3.25). In addition, longer tracks of leftward- or rightward-moving forks were assigned by forkSense, consistent with the expectation of more prolonged BrdU incorporation in the absence of a thymidine chase. Some molecules showed consistently high BrdU probabilities across their entire length, which meant that there was no obvious drop in signal to enable the prediction of fork direction, rendering the interpretation of replication dynamics impossible (e.g. Mol 2). The fork directionality was assigned in other molecules and once again indicated the presence of multiple initiation events in close proximity on single molecules (e.g. Mol 5 and 6). However, the pattern of BrdU incorporation across these molecules was relatively even, with no clear peaks in BrdU signal, which made these data suboptimal for the interpretation of replication dynamics.

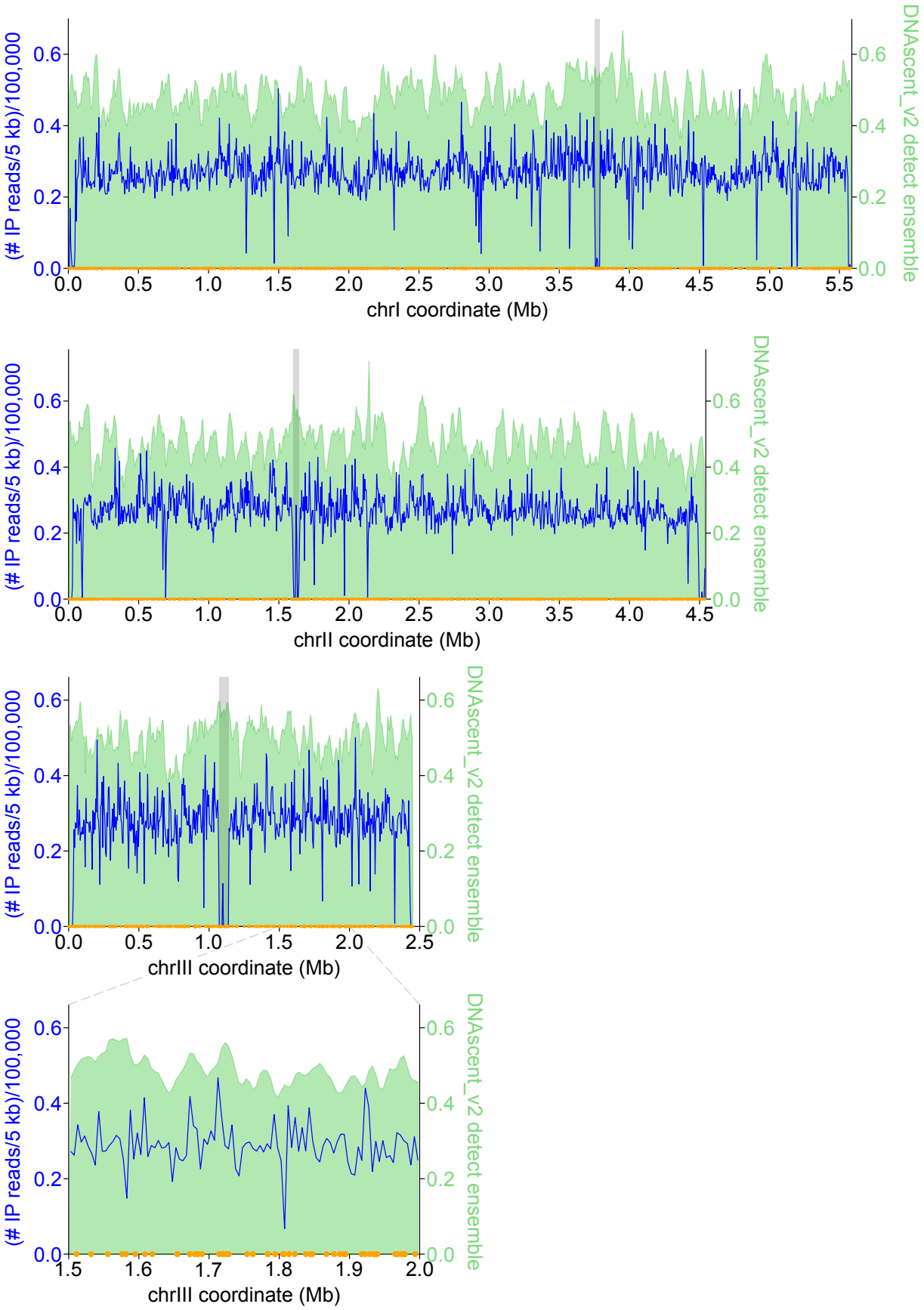


***Figure 3.24**: Visualisation of an ensemble of D-NAscent data (green) with BrdU-IP data from the same time point (blue). The number of IP reads per 5 kb window was divided by 100,000 to enable plotting on a comparable scale to the D-NAscent ensemble data. Orange circles*

along the x axis mark replication initiation sites identified by PU-seq (Daigaku et al. 2015). *Vertical grey bars mark the centromeres.*

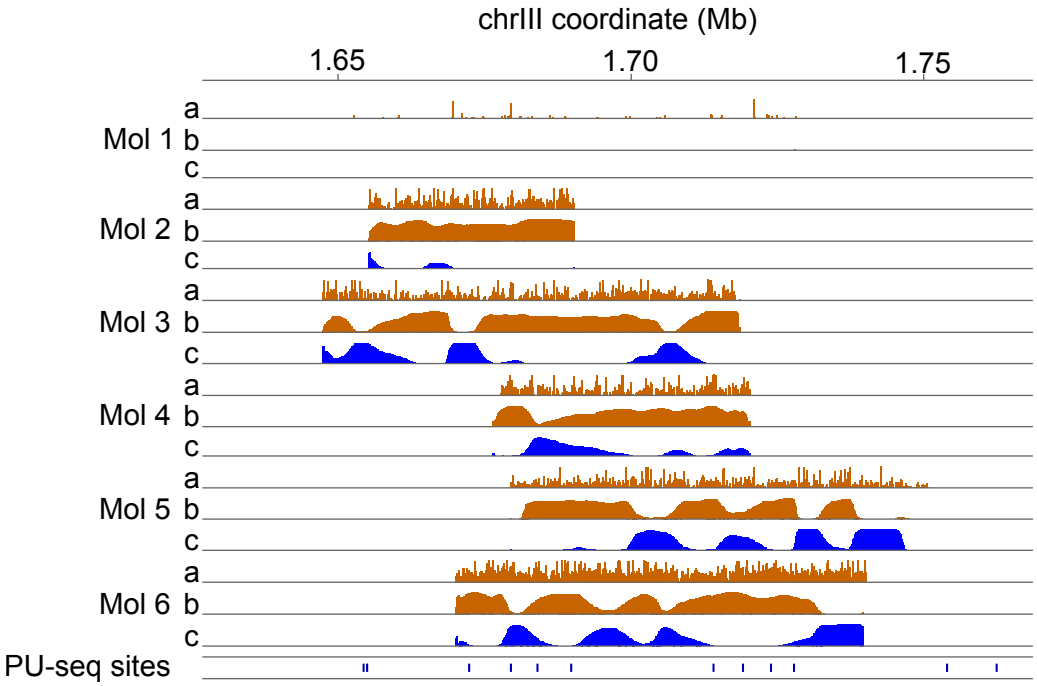


Figure 3.25*: Visualisation of individual molecules aligning to chromosome III, as in Figures 3.16 and 3.19. Each molecule is represented by three tracks: a shows the probability of a given thymidine being a BrdU, b shows the probability of a leftward-moving fork replicating a region, while c shows the probability of synthesis by a rightward-moving fork. The y-axis scale for each track ranges from 0 to 1. Known initiation sites identified by PU-seq* (Daigaku et al. 2015) *are annotated.*

To test whether the initiation sites predicted by forkSense in these data showed a similar level of co-localisation to PU-seq origins as observed in the presence of HU in Figure 3.22, the distance between each such site and its nearest PU-seq origin was plotted (Figure 3.26). Under these conditions, 27,269 origins were called by forkSense, based on its prediction of fork direction in 20,603 single molecules. The median distance between these sites and neighbouring PU-seq origins was 2,764 bp. The distribution of the distances showed a less dramatic bias towards initiation events called in the closest proximity to PU-seq origins than in Figure 3.22. Furthermore, the distribution was very similar to that seen for the same number

of randomly-generated 1,113 bp sites (the median length of the forkSense-predicted origins). This was consistent with less accurate initiation site predictions in these data, and a more random distribution of forkSense-predicted origins, as was expected based on the visualisation of the single molecules in Figure 3.25. Therefore, these experimental conditions were not suitable for D-NAscent analysis of replication dynamics.

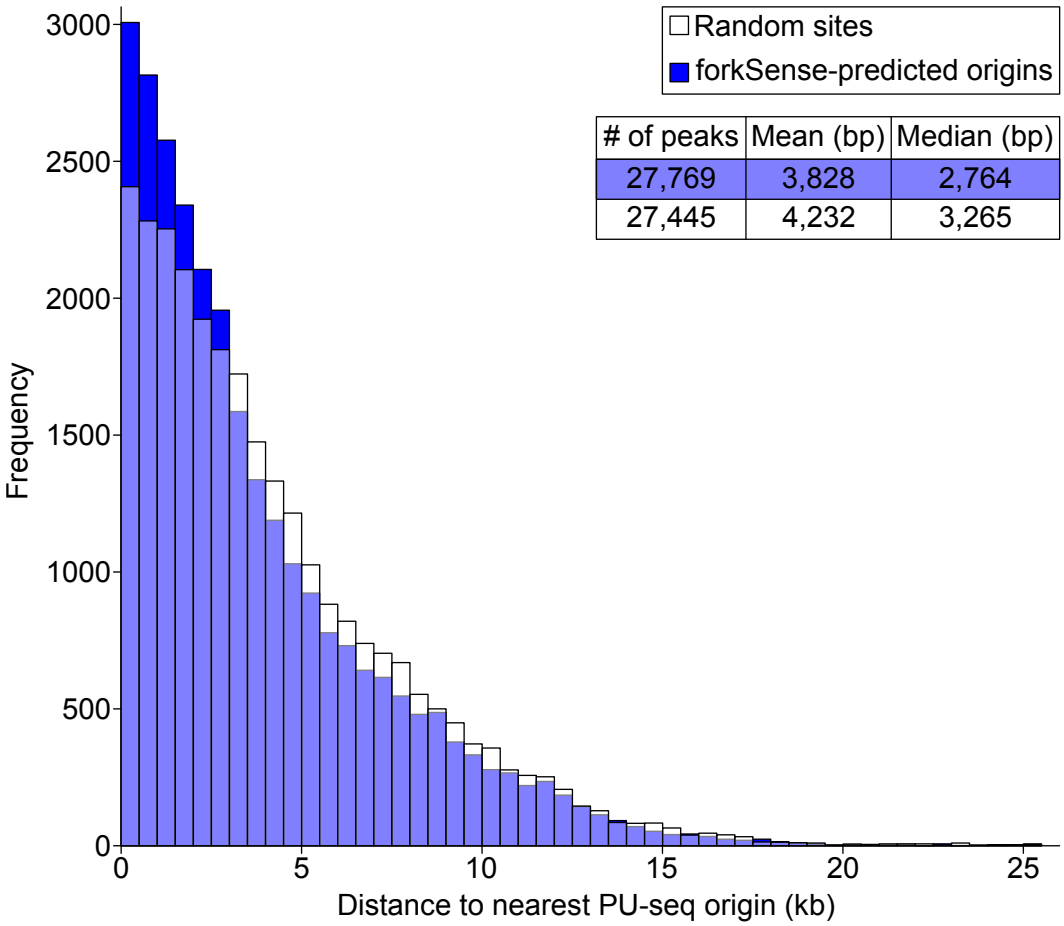


***Figure 3.26**: The distribution of distances between each of the 27,769 forkSense-predicted origins and their nearest PU-seq origins* (Daigaku et al. 2015) *(blue). The median of these distances was 2,764 bp. In white, the distribution of distances between 27,340 randomly-generated 1,113 bp windows and their nearest PU-seq origins was plotted. The median distance between these random windows and the PU-seq origins was 3,265 bp.*

3.6 Summary

To summarise, it was shown that it was possible to detect BrdU incorporated during genome replication in long single molecules from *S. pombe* cells that had proceeded synchronously

through S phase by nanopore sequencing. Therefore, the experimental requirements outlined in section 3.1 had been met. However, the single molecule data indicated that further experimental optimisations were required. The goal of such optimisations would be to promote an increase in BrdU incorporation specifically at regions synthesised at the start of S phase, with lower incorporation elsewhere. This would then result in more clearly-defined BrdU-positive and BrdU-negative regions within the single molecules, facilitating the prediction of fork direction and therefore initiation and termination sites. DNAscent_v2 and the forkSense sub-programme were trained on datasets generated from *S. cerevisiae* (Müller et al. 2019), and so the ability of forkSense to accurately predict fork direction information from a different system (e.g. *S. pombe*) was unknown. Furthermore, based on the data shown here, it seemed that BrdU incorporation did not decrease during S phase in experiments performed without HU, in contrast to what has previously been observed in *S. cerevisiae* (Müller et al. 2019). This suggested that an alternative strategy for the interpretation of replication dynamics may be required.

Chapter 4: Generating and investigating strains for optical mapping of polymerase usage on single molecules

4.1 Introduction

The aim of this project was to develop a novel optical mapping-based approach to track polymerase usage across an entire S phase. The intention was to adapt published population-level methods (Clausen et al. 2015; Koh et al. 2015; Reijns et al. 2015; Daigaku et al. 2015) for analysis at the single molecule level, using Bionano Genomics (BNG) technology.

The approach was based on previous studies that had described protocols to show the contribution of each replicative polymerase in the synthesis of each strand (Clausen et al. 2015; Koh et al. 2015; Reijns et al. 2015; Daigaku et al. 2015). These techniques exploited known mutations in the replicative polymerases that result in increased incorporation of ribonucleotides into nascent DNA during replication. The required polymerase mutations have been identified in both budding and fission yeast (Daigaku et al. 2015; Lujan, Williams, Clausen, Clark, and Kunkel, 2013; Miyabe, Kunkel, and Carr, 2011; Nick McElhinny et al. 2010). Ribonucleotide incorporation by wild-type DNA polymerases during DNA synthesis does occur, albeit relatively infrequently (Joyce, 1997; Nick McElhinny, Kumar, et al. 2010; Nick McElhinny, Watts, et al. 2010). However, ribonucleotides increase the susceptibility of DNA to hydrolysis at the ribose phosphate in the backbone (Williams, Lujan, and Kunkel 2016). Therefore, it is vital that incorporated ribonucleotides do not persist in the DNA, thereby compromising genome stability. Cells have a repair pathway to remove mis-incorporated ribonucleotides, which is dependent on the activity of RNase H2 (Nick McElhinny, Kumar, et al. 2010; Rydberg and Game 2002). This enzyme recognises and cleaves DNA-RNA hybrids to initiate the removal of the ribonucleotide and repair of the DNA. In the above methods, which relied on the presence of ribonucleotides incorporated by the mutant version of a polymerase, the RNase H2 enzyme activity was removed. This led to the accumulation of ribonucleotides in the genome over the course of multiple S phases. DNA isolated from such strains was

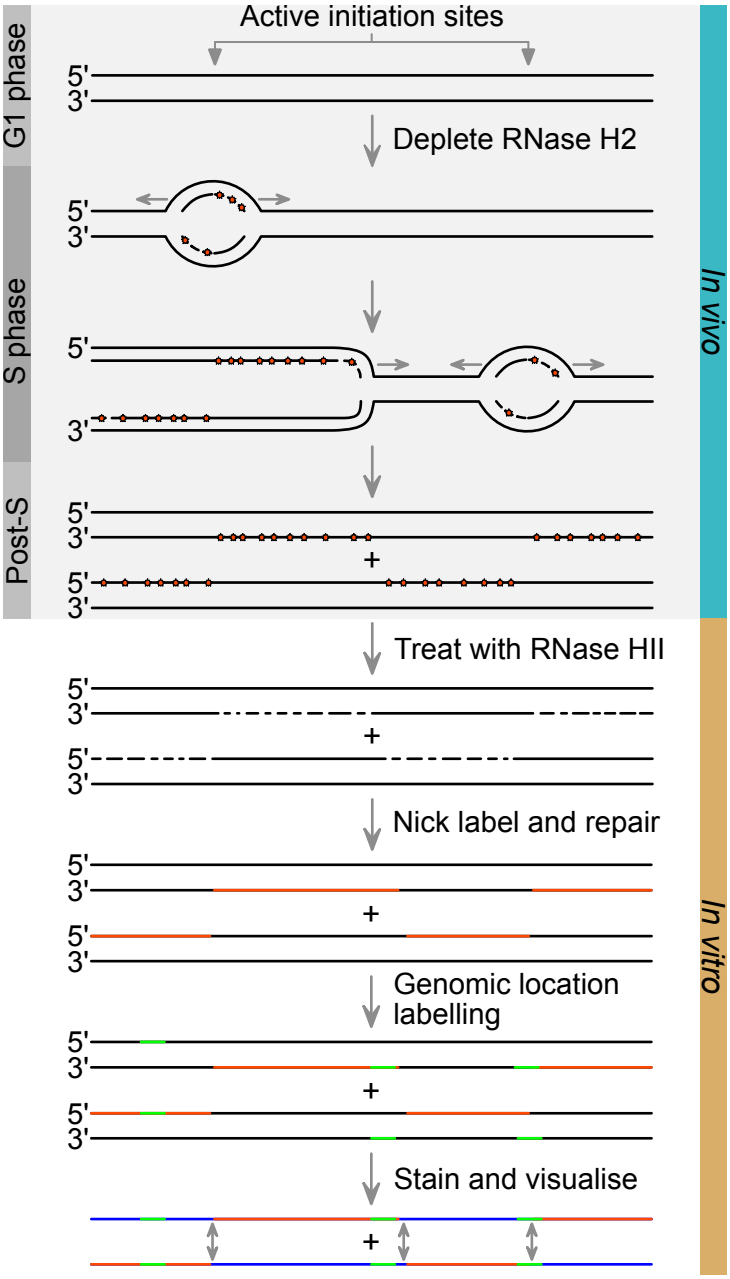


***Figure 4.1**: Outline of the experimental approach to obtain polymerase usage information in ultra-long single molecules by optical mapping. One of the two major replicative polymerases (δ shown here) is mutated such that it has an increased rate of ribonucleotide incorporation compared to wild-type. Additionally, RNase H2 activity is regulated, so that it can be switched off in cells arrested in G1 before entry into S phase. During replication, the ribonucleotides (orange stars) incorporated by the mutant polymerase δ are not removed, and after a single S phase under these conditions, DNA is extracted. The ribose sugars in the DNA backbone are nicked* in vitro *by treatment with RNase HII. The resulting nicks can then be labelled by nick translation (orange), and repaired. After labelling and sealing the nicks, the molecules can then*

undergo genomic location labelling (green) by treatment with DLE-1. Finally, the molecules are stained with YOYO-1 (blue) and can be visualised by Saphyr.

susceptible to nicking at the ribose groups in the presence of an alkali, such as sodium hydroxide. High-throughput short-read sequencing detection of the sites of such nicks was used on DNA extracted from cells expressing one of the replicative polymerase mutants. The resulting data were used to determine which polymerase had been predominantly responsible for the synthesis of each strand in any given genomic region.

In the proposed optical mapping approach (Figure 4.1), DNA from a polymerase mutant strain would be nicked *in vitro* with recombinant RNase HII, rather than NaOH. This would yield a 3' OH group in the DNA backbone upstream of the ribose sugar which is competent for nick translation (Sparks et al. 2012). Conversely, alkaline hydrolysis results in the formation of a cyclic 2',3' phosphate group (Li & Breaker, 1999) which cannot prime strand extension by a DNA polymerase. The resulting nicks could then be labelled using a fluorophore that emits at a different wavelength to the location labelling. This would provide single-molecule polymerase usage information, from which initiation and termination events could be inferred from the transitions between unlabelled to labelled stretches. A key caveat to this proposed method is that, in Irys and Saphyr experiments, each molecule is detected from signal emitted by an intercalating stain (Lam et al. 2012), thus the two strands in dsDNA are indistinguishable and are visualised as one strand. Consequently, it would not be trivial to distinguish initiation and termination events (highlighted by the grey double-headed arrows in the stained molecules in Figure 4.1). Transitions into and out of replication labelling could arise from either event, so this would have to be considered during the data analysis.

An important experimental difference from the published polymerase-usage methods is the requirement for a regulatable system, rather than using an *RNH201* knockout strain. In polymerase mutant strains lacking RNase H2 activity, ribonucleotides accumulate in DNA over

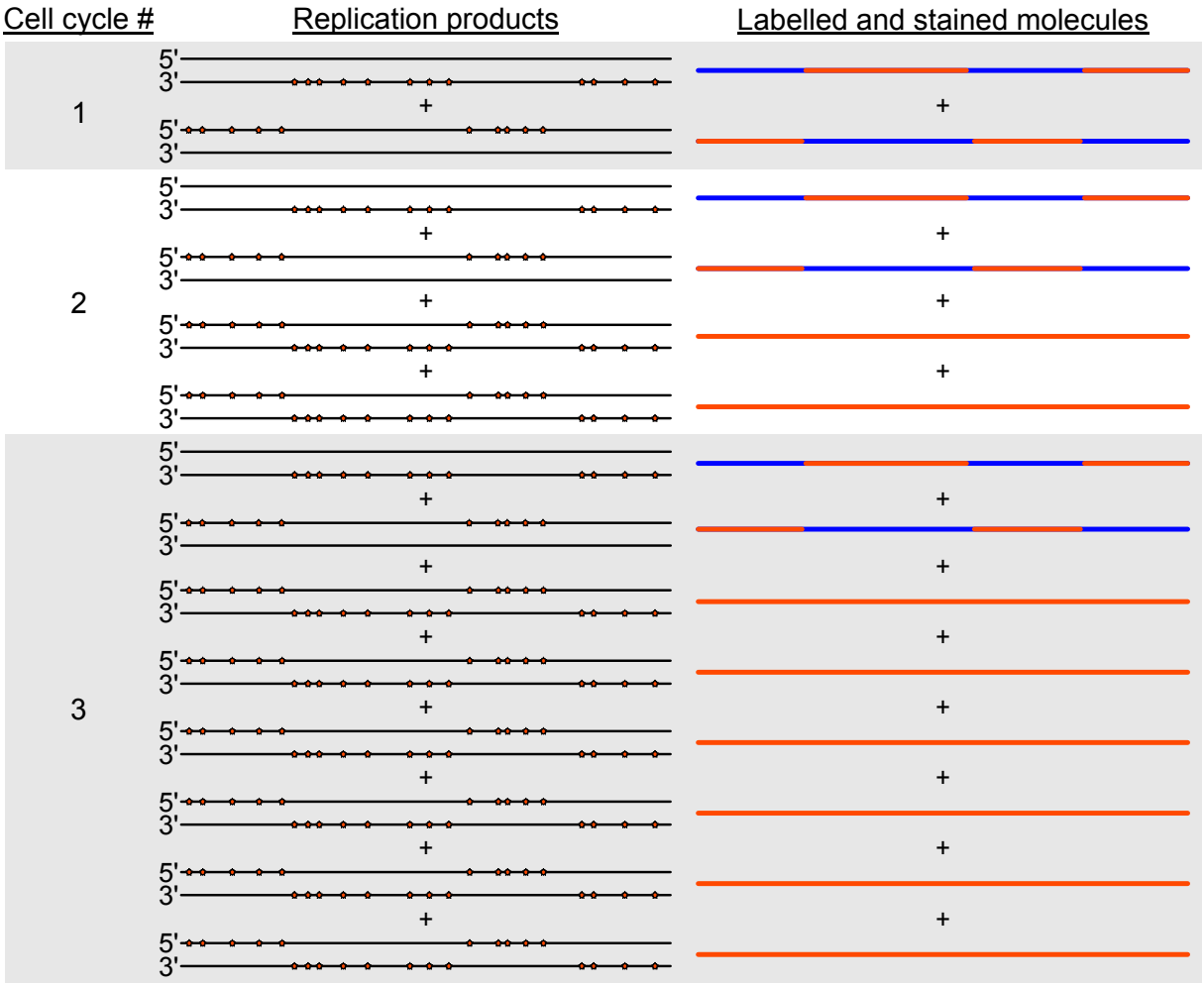


***Figure 4.2**: Optical mapping of genome replication by polymerase-usage requires a system that can be regulated. Here, orange stars in the DNA backbone signify incorporated ribonucleotides. As the cell population goes through multiple cell cycles without RNase H2 activity, ribonucleotides will accumulate on both strands of a DNA molecule in an alternating pattern. In theory, nick-labelling in such molecules could potentially result in label spanning the entire length, and therefore not produce the informative transitions between label and non-label to demarcate the initiation/termination events. Alternatively, the presence of proximal nicks on opposing strands could lead to the formation of frequent DSBs, thereby fragmenting the DNA to such an extent that no useful optical mapping analysis would be possible. Therefore, a requirement for the approach is an experimental system in which RNase H2 activity can be regulated.*

time, as cells undergo successive rounds of replication. This means that, after multiple cell cycles, there is an increasing proportion of molecules with ribonucleotides on both strands of the DNA (Figure 4.2). After the *in vitro* treatments to label the nick sites, the molecules would either be too fragmented or could theoretically display label across the whole molecule when visualised, rather than showing the desired transitions between labelled and unlabelled regions. Thus, an important aspect of establishing this approach would be to switch RNase H2 activity off for a single S phase so that only the nascent DNA would be susceptible to nick-labelling.

4.2 Optical mapping of *S. cerevisiae* and *S. pombe* genomes using Irys

To assess the suitability of BNG optical mapping for yeast DNA, trial experiments were performed on genomic DNA from wild-type *S. cerevisiae* and *S. pombe* strains using the Irys platform. The aim was to confirm that the system could be used for optical mapping in yeast in the absence of replication signal, and to determine what molecule size distribution and genome coverage could be achieved. It would also provide insight into its suitability for the analysis of difficult-to-study regions of the genome, such as the repetitive rDNA array.

Preparation of genomic DNA for an Irys run requires nick, label, repair and stain (NLRS) reactions. First, a nicking endonuclease must be selected based on the frequency of its recognition sites in the reference genome sequence. *In silico* digestion of the sacCer3 reference sequence with Nt.*Bsp*QI yielded a predicted density of 12 labels/100 kb, which was within the optimal range of 8-15 labels/100 kb recommended by BNG. NLRS reactions were performed with Nt.*Bsp*QI on purified high molecular weight wild-type *S. cerevisiae* (T7107) DNA, and the resulting sample was loaded onto an Irys flow cell. For each flow cell, a mosaic image was automatically produced by compiling several backbone-only DNA images from a section of the array. This showed the same section throughout a scan to confirm molecule migration into and out of the nanochannels. Such images were used to check flow cell loading, confirm the presence of molecules in the nanochannels and detect potential nanochannel blockages (Figure 4.3 A). Strong signal in the pillar region early in the scan showed a high

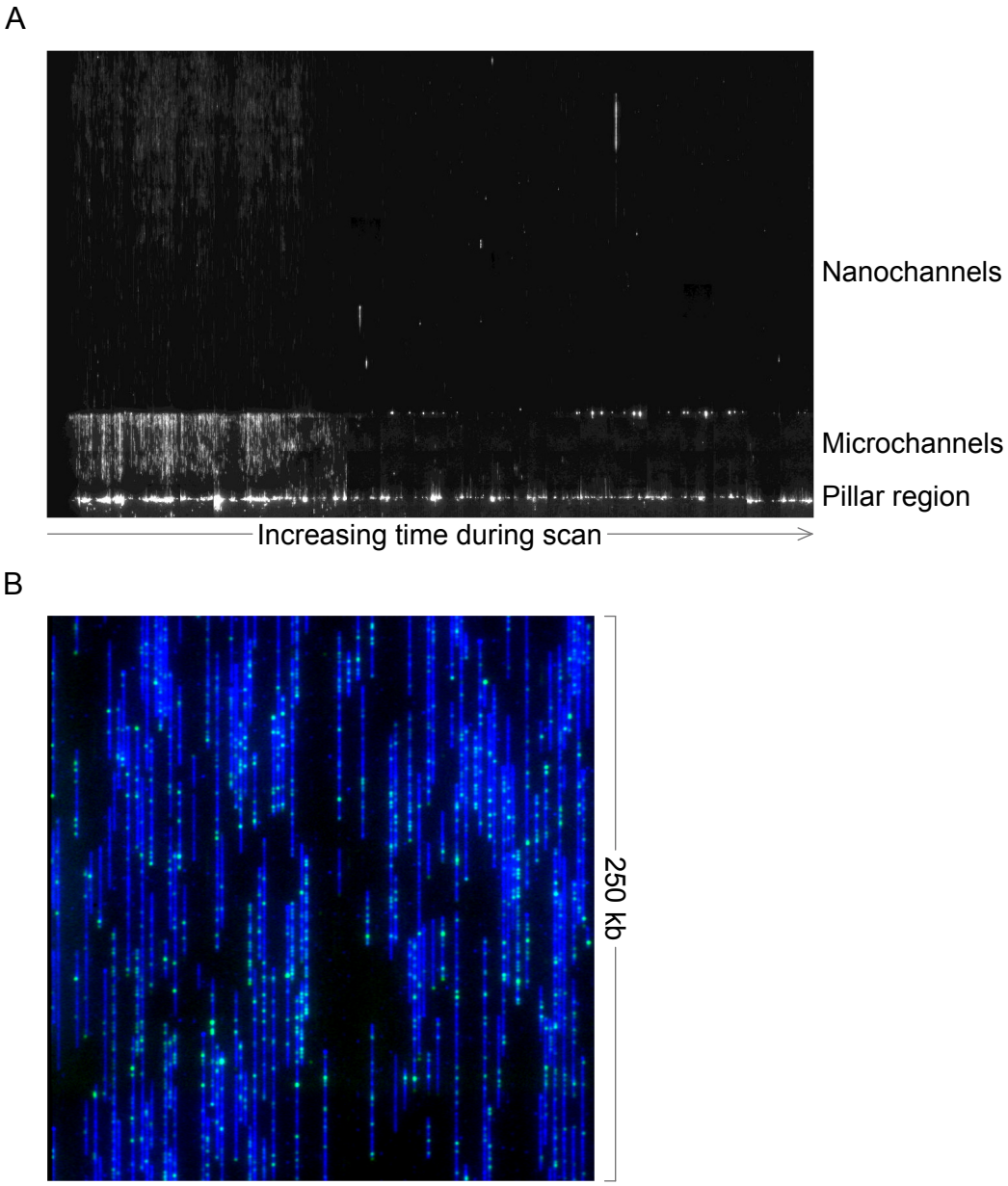


***Figure 4.3**: Visualisation of molecules in nanochannels using Irys. (**A**) A mosaic scan image confirmed migration of molecules through the flow cell. Stained DNA appeared as white signal which revealed molecule progression from the pillar region into the nanochannels. Molecules were detected in nanochannels early in the scan, but not at later time points. The precise timescale of the scan was not recorded. (**B**) An example FOV from the Irys run, with backbone DNA stain shown in blue and nick labels shown in green. Each image spanned approximately 250 kb.*

density of molecules before migration into the microchannels. The strength of the electrical current was such that individual molecules were drawn from the microchannels into the

nanochannels. These individual molecules appeared as fainter signal which travelled out of the field of view as the time increased, consistent with unrestricted flow through the channels. Signal detected in a nanochannel at the later time point was evidence for a single blocked channel. During the Irys run, molecules were visualised at two wavelengths as they migrated through the nanochannels to capture molecule staining and nick labelling information, as shown in blue and green, respectively (Figure 4.3 B). Each image corresponded to a length of approximately 250 kb and adjacent fields of view (FOVs) were automatically appended to enable the analysis of molecules that spanned FOV junctions or multiple FOVs.

An analogous experiment was performed with a wild-type *S. pombe* strain (309), again using Nt.*Bsp*QI as the nicking enzyme, which had a predicted label density of 13 labels/100 kb in this species. The molecule quality results from the two experiments indicated that the total yield of each flow cell provided sufficient data for >5,400x and >6,800x coverage of the *S. cerevisiae* and *S. pombe* genomes, respectively (Table 4.1). When only molecules >250 kb were considered, the genome coverage was >290x and >750x for *S. cerevisiae* and *S. pombe*, respectively. This showed that it was possible to obtain high coverage of the two species' genomes using ultra-long molecules.

Table 4.1*: Summary of molecule quality reports from each Irys run. The values shown represent the yields from an individual flow cell for each experiment. All molecules included in the summary had a label signal-to-noise ratio (SNR) ≥ 3.*

Species	**Approx. genome size (Mb)**	**Number of scans**	**Total data >20 kb (Mb)**	**Mol. N50 (kb)**	**Average coverage by molecules**	
					>250 kb	**>500 kb**
S. cerevisiae	13.6	25	73,915	93	294	66
S. pombe	13.8	30	94,519	103	753	152

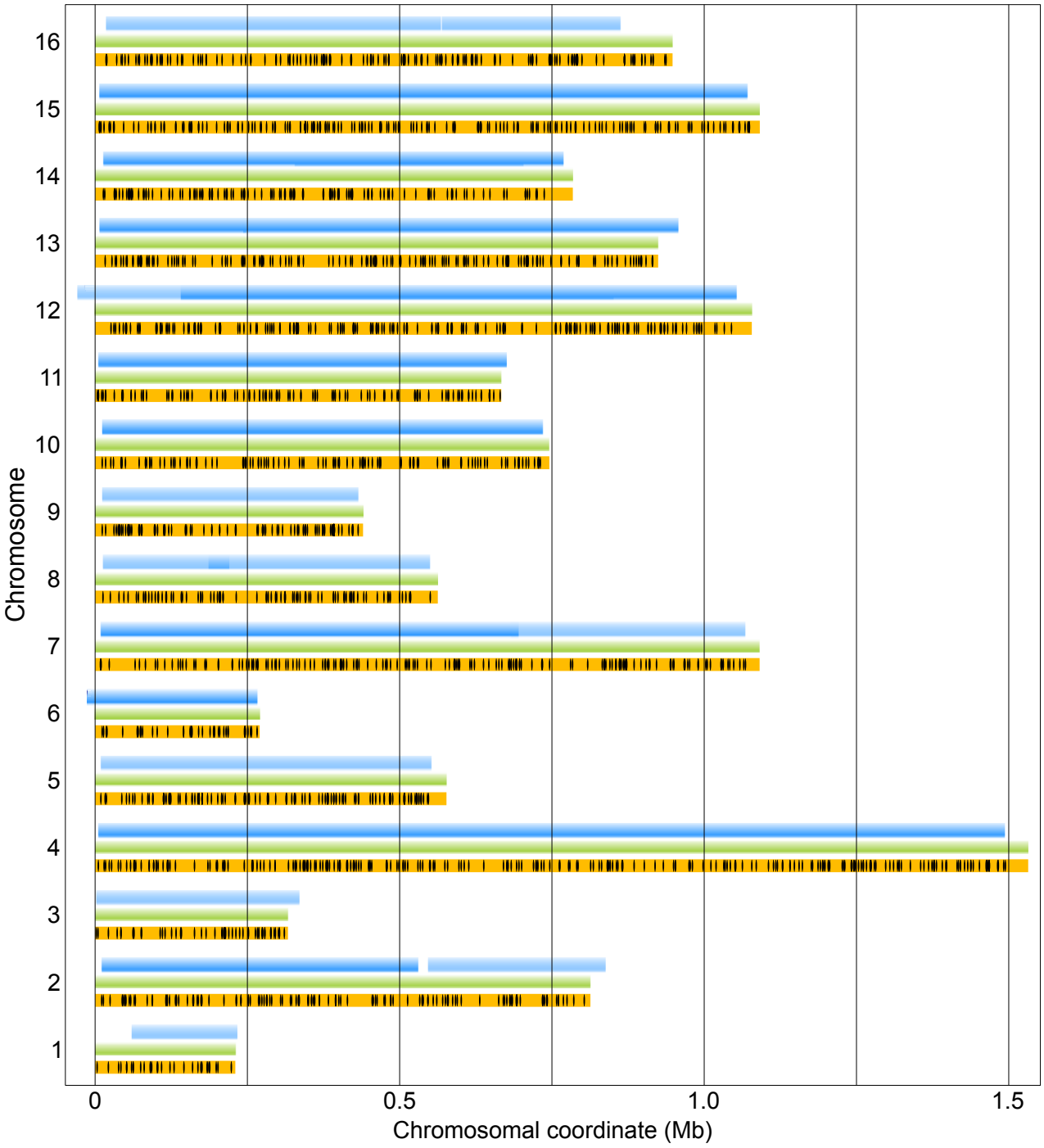


***Figure 4.4**: S. cerevisiae de novo genome assembly using Irys optical mapping data. Chromosomes from the reference sequence are shown in green and the de novo assemblies from the Irys data are shown in blue. The yellow bars represent an in silico digestion of the reference sequence with the predicted nick sites shown in black. The assembly was derived from data in which the genomic DNA was labelled at Nt.BspQI nick sites.*

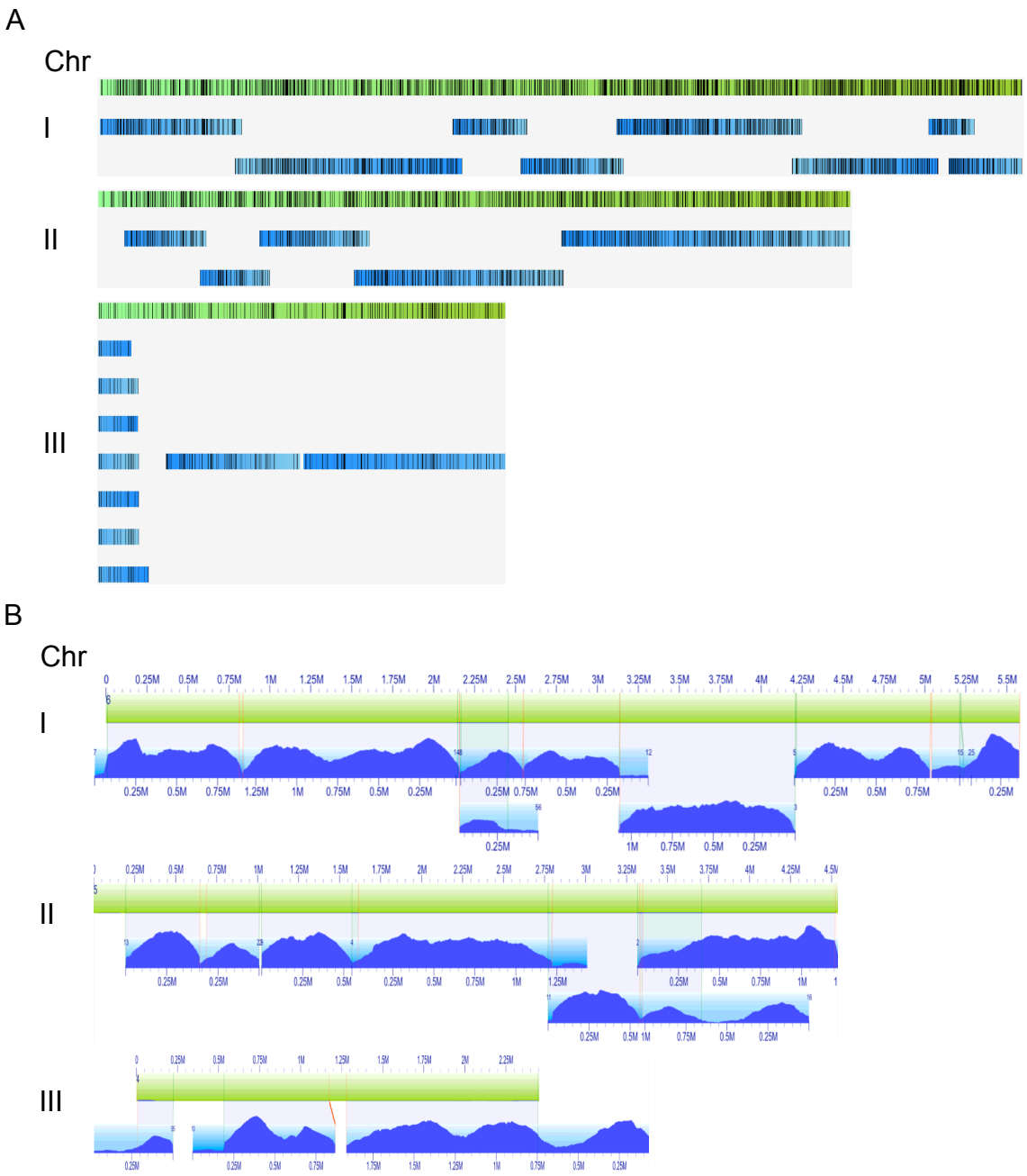


***Figure 4.5**:* De novo S. pombe *genome assembly using Irys. (**A**) The contigs from the* de novo *assembly are shown in blue and the reference chromosomes in green. Each chromosome had multiple contigs forming incomplete chromosome assemblies. (**B**) Visualisation of the molecule coverage (blue) in relation to the reference chromosomes (green). The coverage across the genome was not uniform, with troughs indicating low coverage in several regions.*

The data from these Irys runs were used for a *de novo* assembly of each genome using IrysView software from BNG (Figures 4.4 and 4.5). The assemblies were largely syntenic with reference chromosomes for both species, with the main exceptions being the telomeres. For

S. cerevisiae, the ends of the assembled contigs (blue in Figure 4.4) tended to be truncated when compared with the reference chromosomes (green). In most cases, this was likely due to the lack of Nt.*Bsp*QI recognition sites in the telomeric DNA, which would preclude optical mapping of these regions using this enzyme.

Next, individual contigs from the *S. pombe de novo* assembly were aligned to the reference. (Figure 4.5 A). This showed that chromosome I comprised eight contigs in the Irys assembly (blue), while chromosome II had five contigs and chromosome III had three. The repetition of the first contig aligning to chromosome III was consistent with incomplete assembly of the rDNA array that is located at the 5' end of chromosome III. A similar difficulty in mapping the rDNA locus was observed in the *S. cerevisiae* data (not shown). This suggested that it was not possible to map across these repetitive elements in either dataset and so, while ultra-long molecules could be analysed using this approach, it remained challenging to span certain genomic regions. Finally, coverage maps were used to assess the variations in coverage across the *S. pombe* assembly (Figure 4.5 B). It is likely that the sites of low coverage included "fragile sites" in the Irys data, where adjacent Nt.*Bsp*QI nick sites resulted in DSBs in the DNA, thereby preventing mapping across these sites.

In summary, it was found that Irys technology could produce high-coverage optical mapping data for both *S. cerevisiae* and *S. pombe* genomes. Crucially, coverage remained high when only long molecules were considered. Therefore, if information about polymerase usage could additionally be obtained, Irys could theoretically be used to provide insight into replication in ultra-long single molecules.

4.3 Using gel electrophoresis and Southern blotting to detect increased ribonucleotide incorporation

Specific point mutations in the replicative polymerases increase ribonucleotide mis-incorporation during DNA replication (Daigaku et al. 2015; Lujan et al. 2013; Miyabe, Kunkel, and Carr 2011; Nick McElhinny, Kumar, et al. 2010). These ribonucleotides persist when

RNase H2 activity is removed, thereby rendering the DNA more alkali-sensitive. Thus, NaOH treatment followed by gel electrophoresis can be used to detect increased ribonucleotide incorporation in DNA (Nick McElhinny, Watts, et al. 2010). Alkali-induced nicking at ribose groups and DNA denaturation results in smaller fragment sizes which can be visualised as a downward shift in signal on an agarose gel. This can be coupled with Southern blotting to reveal strand-specific information through the use of probes (Daigaku et al. 2015).

The initial goal was to confirm the detection of higher ribonucleotide incorporation in *rnh201Δ S. pombe* strains with one of the known replicative polymerase mutations (*cdc6-L591G* and *cdc20-M630F*, for polymerase δ and polymerase ε, respectively), by alkali treatment and gel electrophoresis (Figure 4.6 A). All of the strains were kindly provided by Prof. Carr's lab (University of Sussex). DNA from the wild-type and *cdc20-M630F* strains showed indistinguishable fragment size distributions (Figure 4.6 B). In contrast, DNA from the *rnh201Δ* and *cdc6-L591G* strains showed a small downward shift in signal compared to the wild-type, suggesting a slight increase in fragmentation in these samples. When combined with *rnh201Δ*, both polymerase mutations caused a downward shift in the peak of the signal when compared with the parental polymerase mutants. These results were consistent with higher genomic ribonucleotide levels in the strains with both RNase H2 deleted and one of the replicative polymerases mutated, compared to the polymerase mutant strains with active RNase H2.

Southern blots with strand-specific probes were used to further validate the polymerase mutant strains. Nascent DNA shows a characteristic pattern of polymerase usage, with the leading strand synthesis performed mainly by polymerase ε and the lagging strand synthesis by polymerase δ (Figure 4.7 A) (Miyabe, Kunkel, and Carr 2011; Daigaku et al. 2015). Probes targeting either the Watson (W) or the Crick (C) strand adjacent to an active replication origin can show whether the strand was synthesised by either polymerase ε or polymerase δ when analysing DNA from polymerase mutants lacking RNase H2 activity (Daigaku et al. 2015). The strand synthesised by the mutant polymerase in the majority of cells in the population shows a downward shift in signal when compared with the controls. DNA in the gel shown in Figure

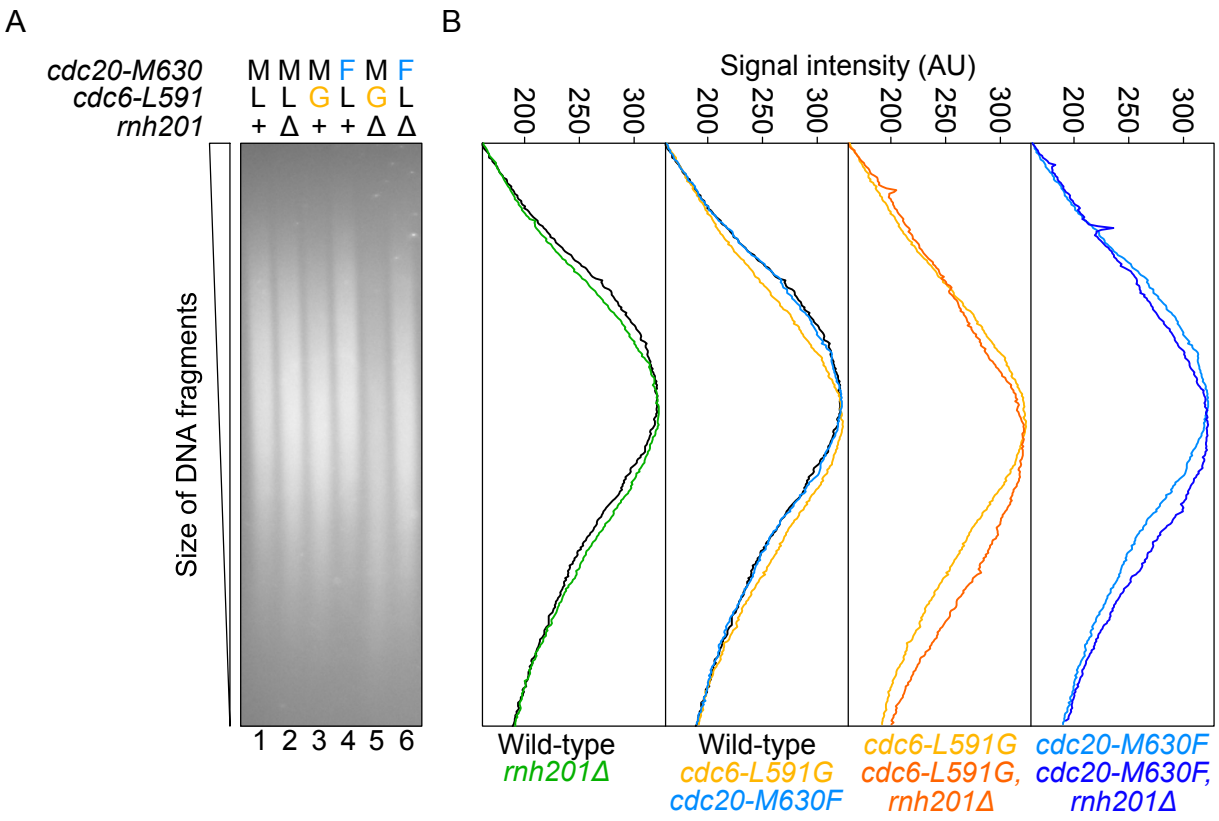


***Figure 4.6**: Detection of alkali-induced fragmentation of DNA in* rnh201Δ *polymerase mutants by denaturing gel electrophoresis. (**A**) Genomic DNA from strains with the indicated genotypes was purified, NaOH-treated and analysed by alkaline gel electrophoresis. (**B**) The signal intensity in each lane was measured. DNA fragment lengths were similar in the wild-type and* cdc20-M630F *mutant. The DNA size distribution showed a slight downward shift compared to wild-type in the* rnh201Δ *and* cdc6-L591G *single mutants. When either polymerase mutation was combined with the RNase H2 knockout, there was a downward shift in signal compared to respective polymerase mutant strains with wild-type RNase H2.*

4.6 was transferred to a membrane and probed. When treated with the W-strand probe, signal from *cdc6-L591G, rnh201Δ* DNA shifted downwards compared to that of the wild-type, the single mutants and the *cdc20-M630F, rnh201Δ* DNA (Figure 4.7 B). This indicated the presence of more ribonucleotides in the W strand, suggesting that it was synthesised by the mutant polymerase δ. Following treatment with the C-strand probe, there was a slight decrease in signal from higher molecular weight DNA and an increase in signal from shorter fragments in the *cdc20-M630F, rnh201Δ* sample when compared to the *cdc20-M630F* DNA (Figure 4.7

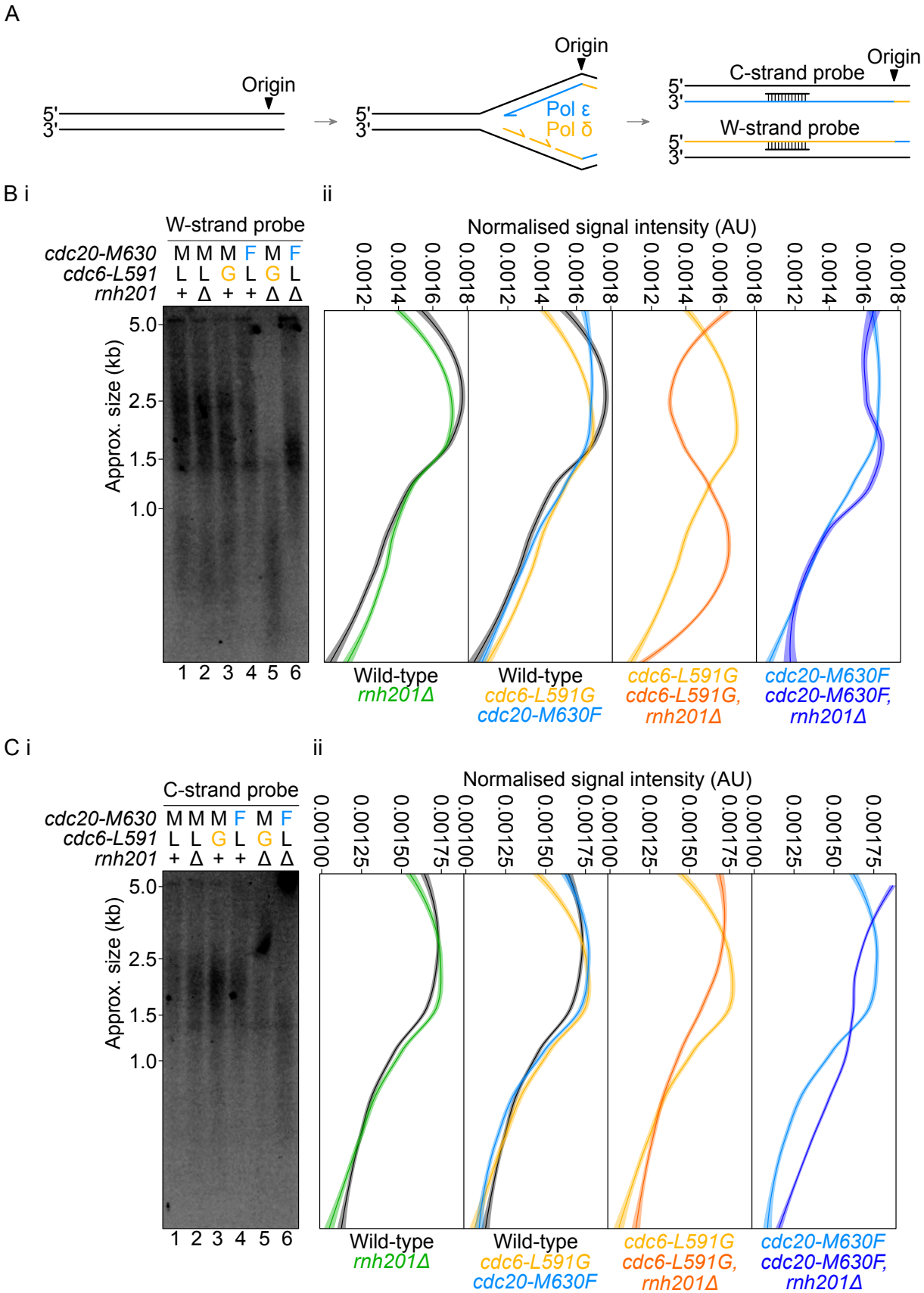


***Figure 4.7**: Detection of strand-specific ribonucleotide incorporation by Southern blot. (**A**) Schematic showing the principle behind strand-specific probes. Nascent DNA adjacent to a highly active origin of replication shows a distinct pattern of polymerase usage. Strand-specific*

probes are complementary to either the W or the C strand. In this experiment, probes targeted a region adjacent to ARS3006 *and* ARS3007 *as in* (Daigaku et al. 2015). *(**Bi**) The DNA in the gel shown in Figure 4.6 was transferred onto a nylon membrane which was then treated with the W-strand probe. (**Bii**) The signal intensity of each lane in (Bi) was measured. There was a marked decrease in fragment sizes in the* cdc6-L591G, rnh201Δ *DNA compared to the wild-type, single mutant and* cdc20-M630F, rnh201Δ *DNA. This was apparent as a change in peak position and decrease in signal in the upper region of the blot. (**Ci**) Treatment with the C-strand probe showed apparently similar fragment lengths in all samples. (**Cii**) Analysis of the signal intensities revealed a slight downward-shift in signal in the* cdc20-M630F, rnh201Δ *sample compared to the* cdc20-M630F *single mutant. Data in signal intensity plots were normalised to the area under the curves and smoothed by the Loess approach with the 95% confidence limits shown in lighter blocks around the smoothed lines.*

C). This was consistent with increased ribonucleotides in the C strand in this double mutant, indicative of Pol ε-dependent synthesis. Therefore, the W strand was largely synthesised by Pol δ, and the C strand by Pol ε in this experiment.

In summary, it was shown that alkali treatment followed by gel electrophoresis could be used to detect the presence of elevated ribonucleotide levels in polymerase mutants lacking RNase H2 activity. Thus, this assay could be used to screen the conditional mutants that would be required for optical mapping of polymerase usage genome-wide. Southern blotting could provide additional information in the form of strand-specificity through the use of appropriate probes.

4.4 Generating and testing *S. pombe* strains for the polymerase-usage-based approach

4.4.1 Regulating RNase H2 activity using a hormone-binding domain

Optical mapping of polymerase usage requires increased ribonucleotide incorporation for a single S phase to obtain signal transitions demarcating sites of changes in polymerase usage

(e.g. initiation and termination sites). Regulation of RNase H2 activity is one potential way to conditionally control increases in genomic ribonucleotide levels. It has previously been shown that attaching a hormone-binding domain (HBD) to a protein in *S. pombe* can regulate its activity through the addition and removal of the hormone β-oestradiol (Bøe et al. 2008; Pai et al. 2009, 2012). In the absence of β-oestradiol, the HBD binds to heat shock protein 90 (Hsp90), which can block protein activity through steric hindrance. In addition, the correct localisation of nuclear proteins can be prohibited, thereby separating HBD-tagged proteins from their substrates and/or binding partners (Pai et al. 2012). Protein activation is achieved through the addition of β-oestradiol, which binds to the HBD, thereby blocking Hsp90 binding. Under these conditions, the tagged protein should no longer be sterically blocked or mis-localised by Hsp90. Thus, it was hypothesised that RNase H2-HBD activity might be inhibited upon removal of β-oestradiol, thereby preventing the removal of ribonucleotides from the genome.

To determine whether the HBD could provide regulation of RNase H2 activity, it was necessary to tag an endogenous RNase H2 gene with this protein. To facilitate this approach, a tagging vector was kindly provided by Prof. Kearsey. This vector (p637) was modified by integration of the *rnh202*$^{+}$ gene upstream of the HBD DNA, producing pEDH5. This construct was linearised and then integrated at the endogenous *rnh202*$^{+}$ locus in a wild-type *S. pombe* strain by homologous recombination (Figure 4.8). The *rnh202*$^{+}$ gene encodes the ancillary B subunit of the RNase H2 heterotrimer. It was targeted in this strategy as it was considered to be less detrimental than tagging the C-terminus of either the catalytic A or ancillary C subunits. Successful transformants were isolated by positive selection on nourseothricin, genotyping PCRs, and Sanger sequencing of the junction between *rnh202* and HBD DNA.

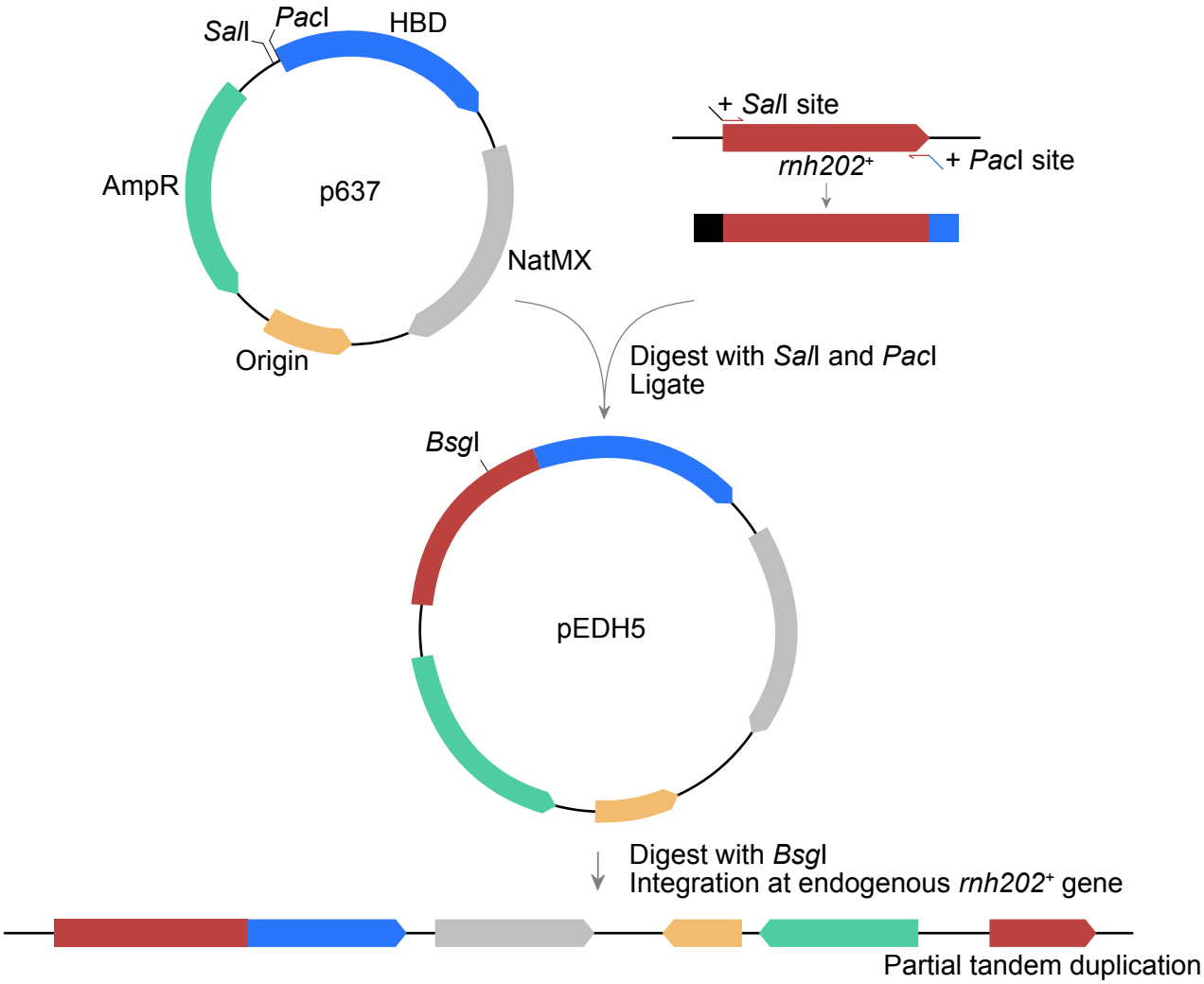
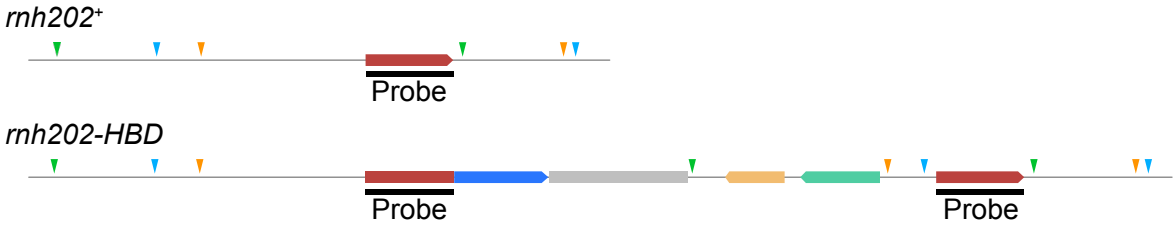


***Figure 4.8**: Strategy for fusing the HBD tag to the 3' end of the* rnh202^{+} *gene. The* rnh202^{+} *gene was amplified by PCR and then cloned into the empty tagging vector (p637) using restriction enzymes* SalI *and* PacI*. The resulting tagging construct (pEDH5) was linearised by* BsgI *digestion and integrated at the endogenous* rnh202^{+} *locus by homologous recombination.*

Two of the nourseothricin-resistant isolates were genotyped at the *rnh202* locus by Southern blotting. Genomic DNA was extracted and digested with three restriction enzymes in separate reactions. The enzymes were chosen to give a different restriction fragment length pattern upon successful integration of the tagging construct, compared to the untagged gene (Figure 4.9 A). An *rnh202*-specific probe enabled analysis of the restriction fragments at this locus. Digestion with either *Eco*RI or *Ssp*I gave clear differences in the transformants compared to the wild-type, with the strongest signal in each case corresponding to bands of the expected sizes (Figure 4.9 B). The *Nde*I digestion showed faint bands at the expected sizes for each of

A



B

rnh202 genotype	Expected fragment sizes (bp)		
	*Eco*RI	*Nde*I	*Ssp*I
+	4116	4253	3680
HBD	6418 3442	7745 2252	6927 2497

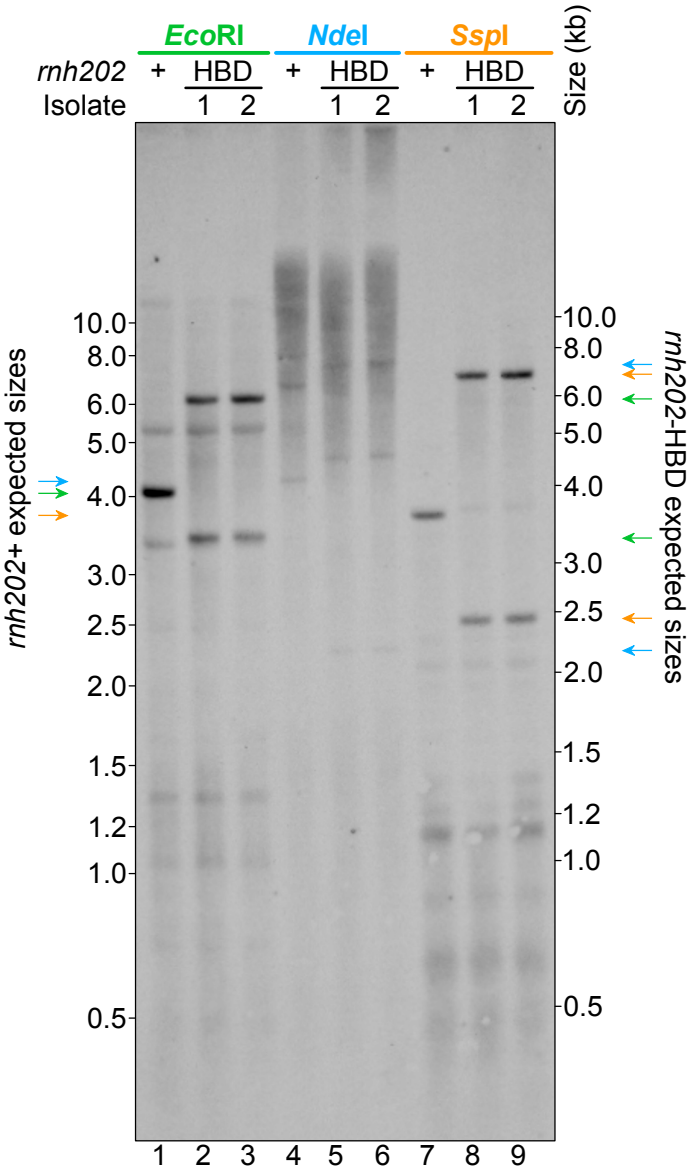


***Figure 4.9**: Southern blot genotyping of* rnh202-HBD *tagging construct. (**A**) Successful integration of the* rnh202-HBD *tagging construct gives rise to a different restriction digest pattern compared to the wild-type locus. Green arrows mark* EcoRI *recognition sites, blue arrows –* NdeI *sites, and orange arrows –* SspI *sites. (**B**) The table indicates the anticipated*

fragment sizes of the wild-type or modified region upon digestion with the indicated enzyme. Two transformants were tested, along with the parental wild-type control. DNA was digested, size separated on a gel, and transferred to a membrane. The membrane was probed with an rnh202-*specific probe. The three arrows to the left of the blot indicate the anticipated fragment sizes following digestion of the wild-type locus, with the arrow colour relating to the colour of each enzyme in the table. The arrows to the right of the blot show the expected sizes of the digestion products of the modified locus. In each lane, the strongest signal corresponds to bands of the expected sizes for the designated genotype.*

the samples, but much of the signal was found in high molecular weight smears. This was consistent with incomplete digestion. The detection of unanticipated faint bands in each lane was likely due to off-target binding of the probe. Overall, the blot confirmed the PCR and sequencing results that the two transformation isolates had the expected genotype of an HBD tag at the 3' end of the *rnh202*$^{+}$ gene.

An *rnh202-HBD* strain was crossed with each of the two polymerase mutants (*cdc6-L591G* and *cdc20*-M630F) to create double mutants in which the conditional regulation of RNase H2 could be assessed. The resulting strains were grown in the presence or absence of β-oestradiol and harvested for DNA isolation. Extracted genomic DNA was fragmented, treated with NaOH, and size-separated by gel electrophoresis (Figure 4.10 A). The control polymerase mutant, *rnh201Δ* strains showed a decrease in fragment sizes compared to strains with the polymerase mutations and wild-type RNase H2 (Figure 4.10 B). A similar downward shift was expected in samples from polymerase mutant, *rnh202-HBD* cultures grown without β-oestradiol if the HBD was successfully regulating RNase H2 activity. However, the fragment sizes of polymerase mutant *rnh202-HBD* strains were similar to those of the single polymerase mutations, irrespective of the presence or absence of β-oestradiol. This suggested that there was no detectable increase in persistent ribonucleotides in the DNA in polymerase mutant, *rnh202-HBD* strains grown without β-oestradiol, and thus no conditional regulation of RNase

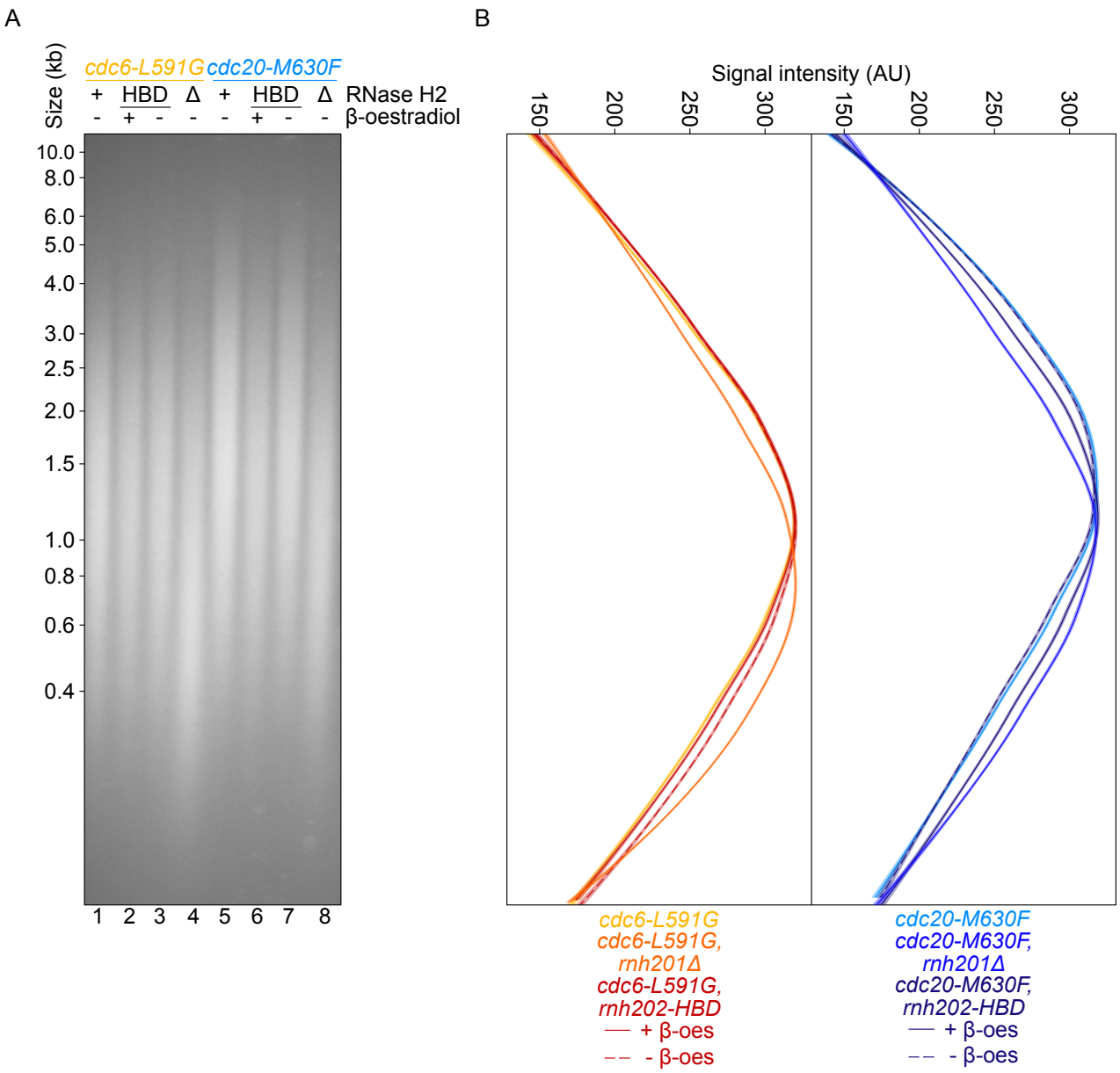


***Figure 4.10**: Assessment of HBD-tagged RNase H2 activity in the presence and absence of β-oestradiol by gel electrophoresis. (**A**) NaOH-treated DNA from the indicated strains was size separated by gel electrophoresis. (**B**) Measurement of the signal intensity of each lane in (A). The data were smoothed by the Loess approach with the 95% confidence limits shown in lighter blocks around the smoothed lines. Growth in the absence of β-oestradiol did not alter the distribution of DNA fragment sizes in either polymerase mutant,* rnh202-HBD *strains when compared to the same strains grown with β-oestradiol. In both cases, the fragment sizes were equivalent to those from control strains in which RNase H2 was fully active. There was a clear decrease in fragment lengths in the* cdc6-L591G, rnh201Δ *and* cdc20-M630F, rnh201Δ *strains which provided a control for lack of RNase H2 activity in each polymerase mutant.*

H2. Later work with the p637 plasmid identified an erroneous deletion that included the *ADH1* terminator downstream of the HBD stop codon. It was not investigated how this impacted on the expression of *rnh202-HBD* but this could have contributed to the result. In the meantime, other methods of RNase H2 regulation were being explored, thus HBD-based control was not pursued further.

4.4.2 Transcriptional regulation of RNase H2 using the nmt81 *promoter*

In *S. pombe*, a common method for controlling gene expression is through the use of the *nmt* (no message in thiamine) promoters (Maundrell 1990; Basi, Schmid, and Maundrell 1993). To explore whether an *nmt*-promoter was able to provide suitable control of RNase H2, a strain in which the *rnh201*$^{+}$ gene was under the control of the *nmt81* promoter (kindly shared by Prof. Carr's lab) was crossed with each polymerase mutant. Haploid strains with both a polymerase mutation and *pnmt81-rnh201*$^{+}$ were isolated. These strains were grown in the presence or absence of thiamine to test whether thiamine addition effectively switched off RNase H2 activity, thereby causing an increase in ribonucleotides in the genome. Purified DNA was treated with NaOH, and size-separated by gel electrophoresis (Figure 4.11). As a control, DNA was treated with H_2O rather than NaOH to assess the integrity of the samples. This showed that the majority of the extracted DNA was of high molecular weight, although there was some fragmentation in samples from cells that had been grown in thiamine (Figure 4.11 A). This downward shift in signal was even more apparent when the DNA had been treated with alkali (Figure 4.11 B). NaOH-treated DNA from the *cdc6-L591G*, *pnmt81-rnh201*$^{+}$ strain grown in the presence of thiamine the showed greater fragmentation than that from the same strain grown the absence of thiamine. A similar thiamine-related increase in fragmentation was seen for the *cdc20-M630F*, *pnmt81-rnh201*$^{+}$ strain. However, given the fragmentation in the H_2O-treated controls, it was unclear whether these results were indicative of an accumulation of ribonucleotides in DNA through repression of RNase H2 expression. It could have been attributed to an RNase H2-independent effect of the thiamine in the culture medium, or a technical step during the DNA extractions. In addition, NaOH-treated DNA from both *pnmt81-rnh201*$^{+}$, polymerase mutants grown in thiamine had smaller fragment sizes than DNA from

either polymerase mutant with *rnh201*$^{+}$ deleted. Since the genomic ribonucleotide levels of the *pnmt81-rnh201*$^{+}$, polymerase mutant samples should not exceed that of the polymerase mutant, *rnh201Δ* strains, this supported the hypothesis that some technical difference during the DNA extraction, or an unrelated thiamine-based effect, contributed to the accumulation of breaks in the DNA.

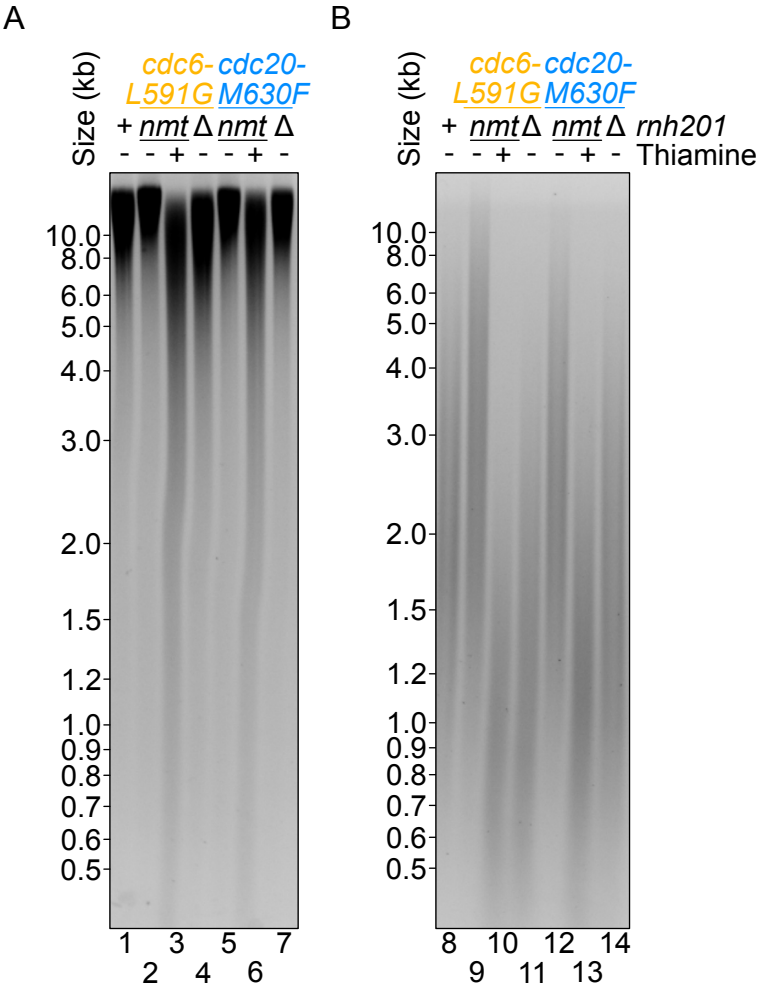
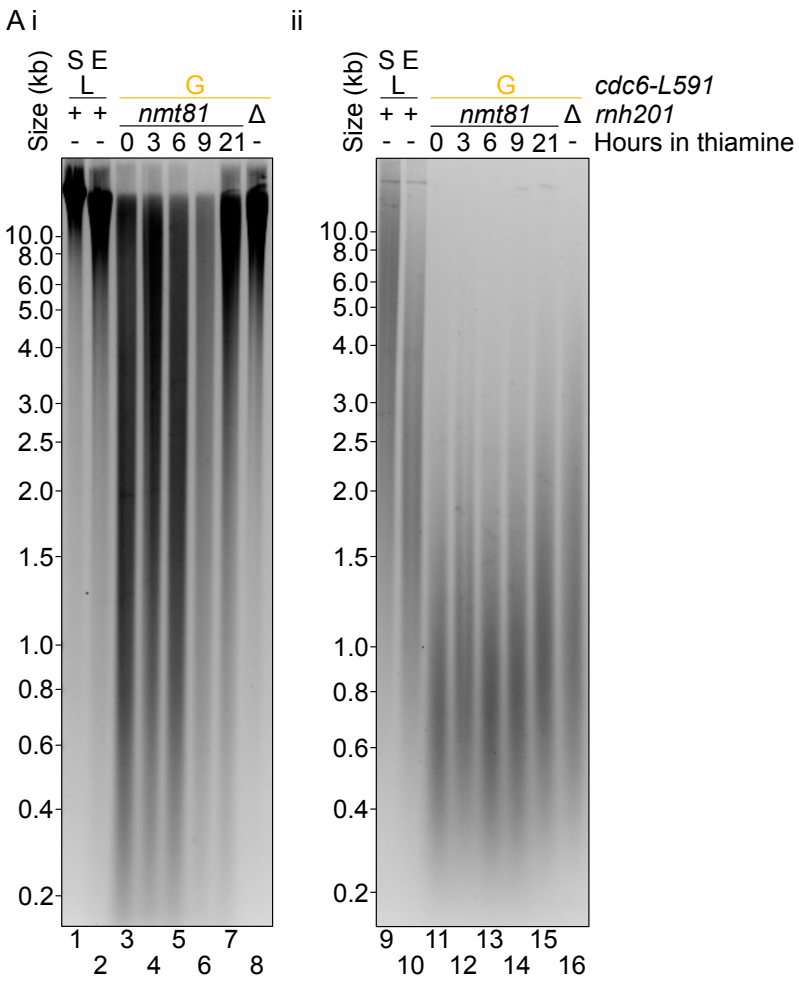


***Figure 4.11**: Assessment of alkali-sensitivity in polymerase mutant strains with the catalytic subunit of RNase H2 under the regulation of the* nmt81 *promoter. (**A**) DNA from the indicated strains was treated with H_2O as a control prior to size separation on a native agarose gel. The majority of the DNA in each of the samples was of high molecular weight, although there was increased fragmentation in samples from cultures treated with thiamine (lanes 3 and 6) relative to the other samples. (**B**) The same DNA as in (A) was treated with NaOH and run on the same gel. DNA from cultures with* pnmt81-rnh201^{+} *and either polymerase mutation grown in thiamine showed the greatest fragmentation of all samples (lanes 10 and 13).*

Time courses of thiamine addition were used to assess whether RNase H2 expression could be controlled in one or more cell cycles. The doubling times of the relevant strains were approximately three hours, so samples were taken at 3 hr intervals to assess whether there was a detectable increase in alkali sensitivity from one cell cycle of thiamine addition up to seven. In previous experiments, high levels of fragmentation had been observed following alkali treatment, even in the wild-type samples (e.g. Figure 4.6). To address whether this indicated the presence of inherent nicks in the DNA of exponentially growing cells, a further control sample was included in this experiment from wild-type cells that had reached stationary phase. This was to assess whether there was less DNA fragmentation than in the exponentially growing wild-type cells. The stationary phase culture DNA did show less fragmentation than the DNA from exponential phase cells in both the H_2O- and the NaOH-treated lanes, although considerable fragmentation was still observed following NaOH treatment (Figure 4.12 A). Given that these cells were not actively cycling, and therefore likely to have fewer sites of DNA damage due to normal DNA metabolism and replication stress, the cause of this extent of DNA fragmentation was unclear. During the time course of thiamine addition with the *cdc6-L591G, pnmt81-rnh201*$^+$ strain, there was no apparent increase in fragmentation of the NaOH-treated samples. Indeed, as the length of time in thiamine increased, the DNA fragment sizes also increased, which was the inverse of the anticipated pattern of decreases in fragment sizes due to ribonucleotides accumulating over time in thiamine. However, since the H_2O-treated samples showed a high level of fragmentation, the results were confounded by unexplained differences in the samples that led to increased fragmentation without alkali treatment of the DNA. The same time course of thiamine addition was performed with the *cdc20-M630F, pnmt81-rnh201*$^+$ strain (Figure 4.12 B). The low fragmentation of the H_2O-treated samples was consistent with improved DNA integrity compared to the *cdc6-L591G, pnmt81-rnh201*$^+$ samples. However, all lanes appeared to show a similar fragment size distribution following NaOH treatment (Figure 4.12 Bii), including the controls. This suggested that each sample was equally susceptible to alkali-induced nicking regardless of the *rnh201*$^+$ promoter and length of time in thiamine.


A i
ii
Size (kb)
S E
L
+ +
- -
G
nmt81
Δ
0 3 6 9 21 -
cdc6-L591
rnh201
Hours in thiamine
10.0
8.0
6.0
5.0
4.0
3.0
2.5
2.0
1.5
1.0
0.8
0.6
0.4
0.2
1 3 5 7
2 4 6 8
9 11 13 15
10 12 14 16

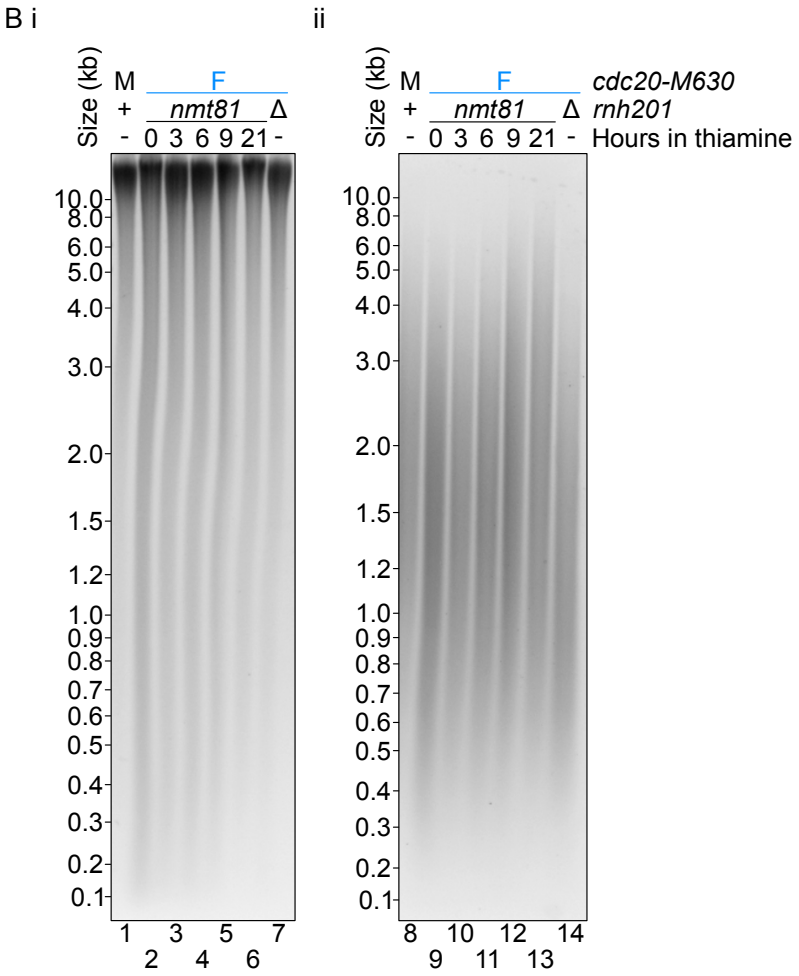

B i
ii
Size (kb)
M
+
-
F
nmt81
Δ
0 3 6 9 21 -
cdc20-M630
rnh201
Hours in thiamine
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.2
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
1 3 5 7
2 4 6
8 10 12 14
9 11 13

***Figure 4.12**: Analysis of ribonucleotide levels following time courses of thiamine addition in* pnmt81-rnh201^{+} *polymerase mutant strains by native gel electrophoresis. (**A**) A* cdc6-L591G, pnmt81-rnh201^{+} *strain was maintained in exponential growth in the presence of thiamine. An additional wild-type stationary phase (S) control was included (lanes 1 and 9) to compare with DNA from exponentially growing wild-type cells (E, lanes 2 and 10). At the indicated times, cells were harvested for DNA extraction and the DNA was treated with either (i) H_2O or (ii) NaOH and run on a native agarose gel. Several H_2O-treated samples showed substantial fragmentation (lanes 3-7). The S sample showed less alkali-induced fragmentation than the E sample. Molecule lengths remained similar or slightly increased following NaOH treatment as the length of time in thiamine increased. (**B**) As in (A), but with a* cdc20-M630F, pnmt81-rnh201^{+} *strain. H_2O-treated samples showed minimal fragmentation, with the majority of signal corresponding to molecules >10 kb. NaOH-treatment resulted in equivalent molecule size distributions in all samples.*

In summary, preliminary evidence indicated that the *nmt81* promoter was not suited to achieve the level of regulation of RNase H2 activity required for the optical mapping of polymerase usage. In addition, the regulation in this system would have been at the transcriptional rather than the protein level, and therefore would not be as rapid as methods of protein-based control. Thus, concurrent work towards the generation of polymerase-usage *S. cerevisiae* strains was pursued.

4.5 Generating polymerase-usage-based strains in *S. cerevisiae*

Replicative polymerase mutations that reduce selectivity against ribonucleotides have been identified in *S. cerevisiae* (Lujan et al. 2013; Nick McElhinny, Kumar, et al. 2010), as well as *S. pombe* (Miyabe, Kunkel, and Carr 2011; Daigaku et al. 2015). Subsequent work focussed on obtaining polymerase-usage strains in *S. cerevisiae* using these known mutations and RNase H2 or mutant polymerase regulation.

Wild-type polymerase ε has both a higher background ribonucleotide incorporation rate than wild-type polymerase δ, as well as a higher rate in its mutant form as stated in (Reijns et al. 2015). Thus, efforts were focussed towards generating a mutant polymerase ε strain, which was expected to have a higher signal-to-noise ratio than the polymerase δ mutant. The gene encoding the wild-type polymerase ε catalytic subunit (*POL2*) was targeted by CRISPR-Cas9 to introduce the M644G mutation. Successful transformants were identified by allele-specific PCRs and Sanger sequencing. The resulting *pol2-M644G* strain could be crossed to strains with a regulatable RNase H2 to produce polymerase-usage strains for optical mapping.

4.5.1 Characterising the effect of RNH201-3xminiAID *on the growth of a Pol ε mutant*

The auxin-inducible degron (AID) system was identified in plants, where auxin hormones are involved in the regulation of certain transcription repressors (Gray et al. 1999). AID-tagged proteins are regulated through ubiquitin-mediated proteasomal degradation in the presence of auxin (Figure 4.13). This has been shown to function in non-plant cells that have been modified to express *TIR1*, including *S. cerevisiae* (Nishimura et al. 2009; Nishimura and Kanemaki 2014). Tir1 mediates the auxin-dependent interaction between the AID protein and the SCF ubiquitin ligase complex (Gray et al. 1999). An E2 ubiquitin ligase is recruited by the SCF complex, and polyubiquitinates the AID protein, targeting it for proteasomal degradation. Therefore, the levels of an AID-tagged protein can be controlled through the addition and

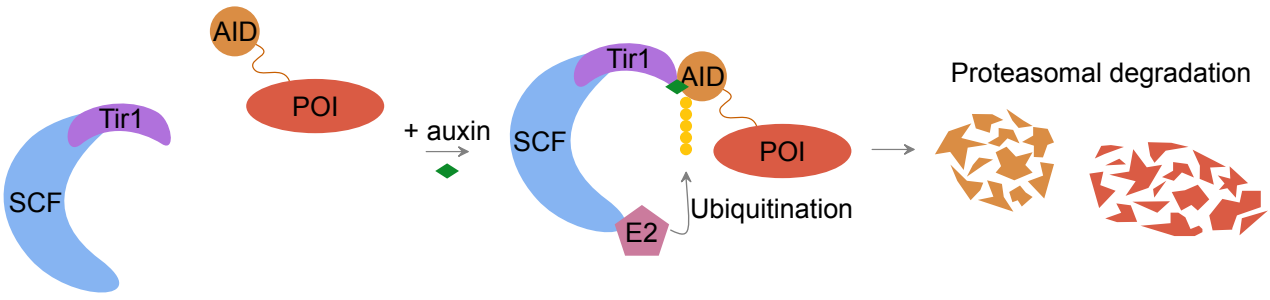


***Figure 4.13**: Schematic showing the regulation of AID-tagged proteins. A protein of interest (POI) is tagged with the degron. Tir1 interacts with the SCF complex, thereby allowing recruitment of the AID protein in the presence of auxin. Poly-ubiquitination of AID by an E2 ligase marks it for degradation by the proteasome, along with the attached POI.*

removal of auxin. Thus, an AID tag on RNase H2 might be suitable to control its activity.

An *S. cerevisiae* strain with 3xminiAID at the 3' end of *RNH201* and constitutively expressed Tir1 (*pADH1-Os.TIR1*) was crossed with a *pol2-M644G* strain. Sporulation and tetrad dissection allowed isolation of strains with all three alleles of interest: *pol2-M644G*, *pADH1-Os.TIR1*, and *RNH201-3xminiAID*. However, the sister spores gave rise to inconsistent colony sizes within a tetrad, with one or two of the spores producing a smaller colony than the others (data not shown). This suggested a growth defect in strains of a certain genotype. Growth on medium lacking uracil was used to identify strains with *URA3-pADH1-Os.TIR1*. Six tetrads that showed 2:2 segregation under this selection were selected for further genotyping by PCR to assess the *POL2* and *RNH201-3xminiAID* genotypes of the *pADH1-Os.TIR1* strains. Of the twelve colonies tested, those with the *pol2-M644G, RNH201-3xminiAID* genotype produced smaller colonies than their sisters. To determine whether the apparent growth defect was related to the *pol2-M644G, pADH1-TIR1, RNH201-3xminiAID* genotype, the doubling times of two isolates with this genotype were measured in liquid culture in the absence of auxin (Table 4.2). A spot assay was also performed to assess growth over a longer period (Figure 4.14). In

***Table 4.2**: Doubling time measurements to assess growth of* pol2-M644G, pADH1-Os.TIR1, RNH201-3xminiAID *strains compared to parental strains in YPAD medium. The parental* pADH1-Os.TIR1, RNH201-3xminiAID *strain had the shortest doubling time, while the two strains isolated with the* pol2-M644G *mutation had a slightly longer time (by 6 min). Two isolates with the* pol2-M644G, pADH1-TIR1, RNH201-3xminiAID *genotype showed an increase in doubling time by more than a third compared to either parental strain.*

Genotype	*RNH201-3xminiAID, pADH1-Os.TIR1*	*pol2-M644G* (Isolate 1)	*pol2-M644G* (Isolate 2)	*RNH201-3xminiAID, pADH1-Os.TIR1, pol2-M644G* (Isolate 1)	*RNH201-3xminiAID, pADH1-Os.TIR1, pol2-M644G* (Isolate 2)
Doubling time (min)	78	84	84	116	116

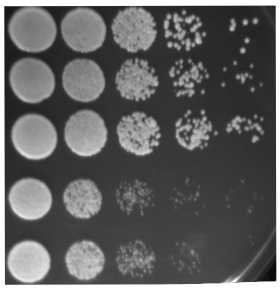


***Figure 4.14**: Spot assay to assess growth of* pol2-M644G, pADH1-Os.TIR1, RNH201-3xminiAID *strains. The same cultures as those used for the measurements in Table 4.2 were diluted in a tenfold series for a spot assay on YPAD medium without addition of auxin. The plate was incubated at 30 °C for one day before imaging. The* pol2-M644G, pADH1-Os.TIR1, RNH201-3xminiAID *strains showed less growth than the parental strains.*

both cases, the parental strains were used as controls. The measured doubling times of the *pol2-M644G, pADH1-TIR1, RNH201-3xminiAID* isolates were 1.4-1.5 times that of the parental strains (Table 4.2) and individual colonies were smaller in the spot assay (Figure 4.14). Both these results suggested an auxin-independent growth defect associated with the *pol2-M644G, pADH1-TIR1, RNH201-3xminiAID* genotype.

To determine whether leakiness in the AID system caused the growth defect in *pol2-M644G, RNH201-3xminiAID, pADH1-Os.TIR1* strains, a strain with a galactose-inducible *TIR1* gene was generated. This was then modified to incorporate the 3xminiAID tag at the 3' end of the *RNH201* gene. After confirming the *RNH201-3xminiAID-KanMX* genotype of two isolates by G418 selection, PCR and Sanger sequencing, mating with a *pol2-M644G* strain was used to introduce this mutation into the resulting diploid. Following sporulation and dissection, the majority of tetrads again showed size disparity between the colonies arising from sister spores, with one or two of the colonies frequently showing less growth than others in the tetrad (Figure 4.15).

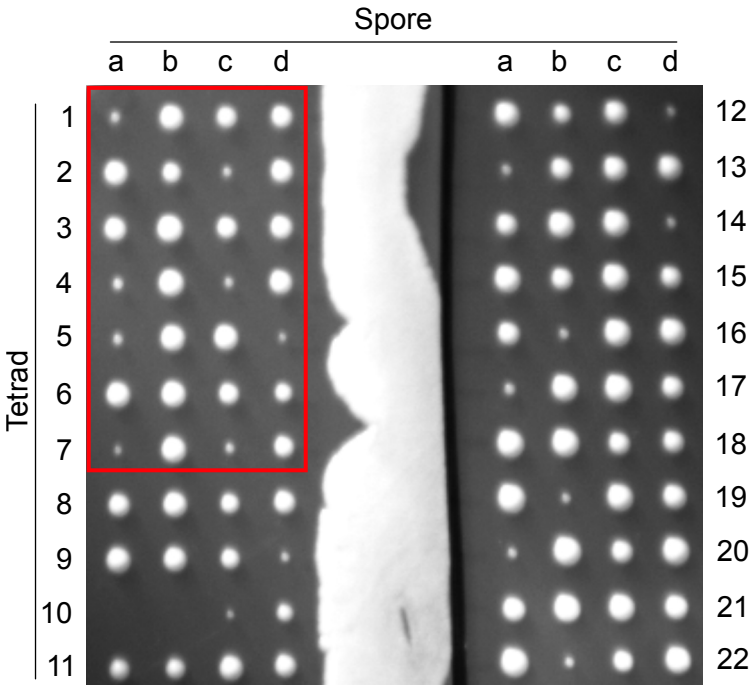
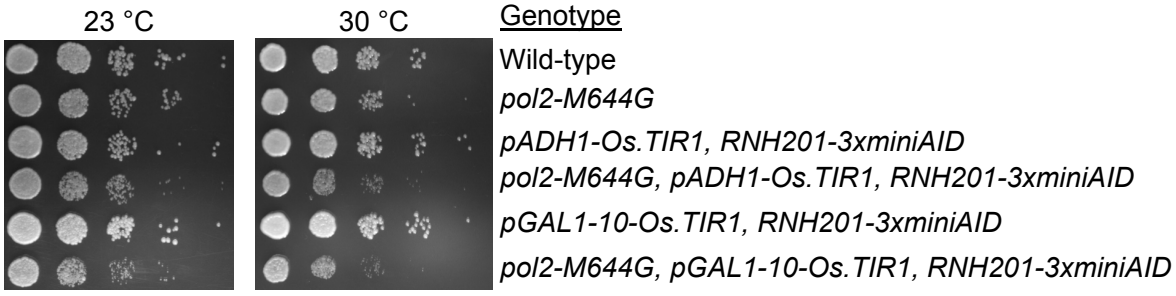


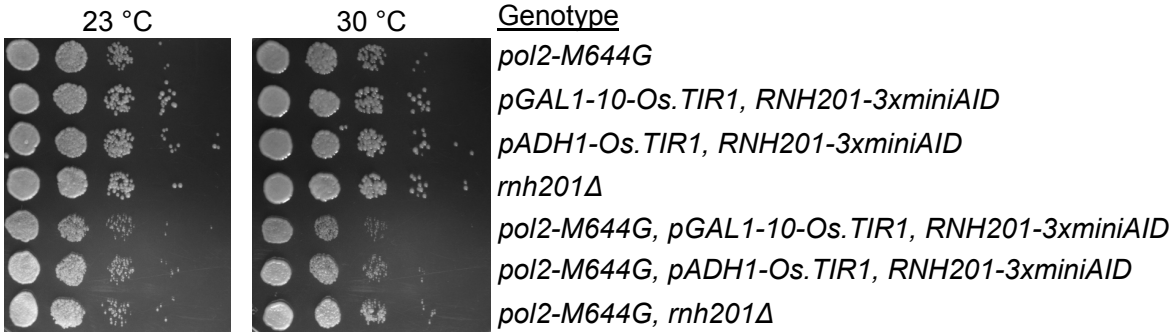
Figure 4.15*: Tetrads from a sporulated diploid heterozygous for* pol2-M644G, pGAL-Os.TIR1, *and* RNH201-3xminiAID*. The tetrads were dissected onto YPAD (containing dextrose, which is repressive for* pGAL-Os.TIR1*). Out of the 22 tetrads, 21 had all four spores grow into colonies. Of these 21 tetrads, seven had four colonies of roughly equal size, 11 had three colonies of roughly equal size with one much smaller colony, and three had two colonies of roughly equal size with two much smaller colonies. Tetrads 1-7 highlighted by the red box were taken forward for genotyping assays.*

To assess which genotype(s) was associated with the growth defect, seven tetrads were tested for the three genotypes of interest (*pol2-M644G, pGAL-Os.TIR1,* and *RNH201-3xminiAID*). This revealed that all of the small colonies in these seven tetrads had the *pol2-M644G, RNH201-3xminiAID* genotype, either with or without *pGAL-Os.TIR1*. This supported the hypothesis that the growth defect was independent of a functioning AID system. To confirm this, spot assays were used to compare the growth of strains with different genotypes on dextrose-containing medium repressive for *pGAL-Os.TIR1*, and lacking auxin (Figure 4.16). It was found that *pol2-M644G, RNH201-3xminiAID* strains with either constitutive *pADH1-TIR1* or repressed *pGAL1-TIR1* had a growth defect compared to wild-type and parental strains (Figure 4.16 A). This defect was comparable to that seen in a *pol2-M644G, rnh201Δ* strain

A



B



C

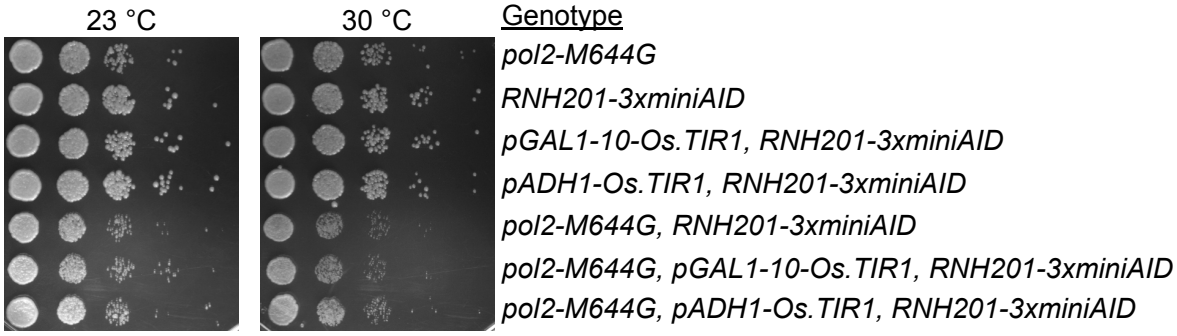


Figure 4.16*: Spot assays to confirm growth defects in* pol2-M644G, RNH201-3xminiAID *strains independently of auxin-induced degradation. Ten strains with the indicated genotypes were spotted onto YPAD without auxin in ten-fold serial dilutions. Each plate was produced in duplicate to allow incubation at 30 °C and 23 °C, to test whether growth at a lower temperature alleviated the defect. Plates incubated at 30 °C were imaged after one day of growth, while plates incubated at 23 °C were imaged after two days. (**A**) A wild-type strain was compared to three parental strains (a* pol2-M644G *strain, a* pADH1-Os.TIR1, RNH201-3xminiAID *strain, and a* pGAL1-Os.TIR1, RNH201-3xminiAID *strain) and two triple-mutant strains (*pol2-M644G, RNH201-3xminiAID *with* pADH1- *or* pGAL-Os.TIR1*). The parental strains showed a mild (*pol2-M644G*) or no growth defect compared with the wild-type, but both triple-mutant strains showed a more severe growth defect at both temperatures. (**B**) The same parental and triple-mutant strains as in (A) were compared to a* pol2-M644G, rnh201Δ *strain. The two triple-mutant strains*

showed an equivalent growth defect to the pol2-M644G, rnh201Δ *strain. (**C**) The same parental and triple-mutant strains as in (A) were compared to a* pol2-M644G, RNH201-3xminiAID *strain with no* TIR1 *gene. This strain showed an equivalent defect to the two triple-mutant strains with either the constitutive or inducible* TIR1 *genes.*

(Figure 4.16 B), and was observed in a strain without a *TIR1* gene (Figure 4.16 C). Therefore, it was concluded that the presence of a C-terminal 3xminiAID tag on the RNase H2 catalytic subunit interfered with RNase H2 activity, thereby causing the accumulation of excess ribonucleotides in the genome in the *pol2-M644G* mutant. Such ribonucleotides would be detrimental to genome stability and therefore were a probable cause of the observed growth defect. This was consistent with observations in *S. pombe* (Prof. Carr lab, personal communication) and with expectations from the literature where C-terminal truncations of the human H2A subunit abrogated complex formation and activity (Reijns et al. 2011; Figiel et al. 2011).

Based on the growth defect observed in *pol2-M644G, RNH201-3xminiAID* strains, it was concluded that the 3xminiAID tag on RNase H2A did not enable the functional regulation of RNase H2 activity. Rather, it appeared to inhibit RNase H2 activity independently of auxin-induced degradation and therefore was not suitable for the optical mapping of polymerase usage.

4.5.2 Characterising the effect of RNH202-AID *on the growth of a Pol ε mutant*

In contrast to the detrimental effects of truncating the human RNase H2A C-terminus (Reijns et al. 2011; Figiel et al. 2011), deletion of the C-terminal region of human RNase H2B did not inhibit enzyme activity (Figiel et al. 2011). Therefore, it was hypothesised that an AID tag on *RNH202* (encoding RNase H2B in *S. cerevisiae*) would be less likely to cause the auxin-independent growth defect seen in the *pol2-M644G, RNH201-3xminiAID* strains above. An *RNH202-AID-9MYC, pADH1-Os.TIR1* strain (generously shared by Dr. Smith's lab) was crossed with a *pol2-M644G* strain and the resulting diploid was sporulated. Following tetrad

dissection, *RNH202-AID-9MYC, pADH1-Os.TIR1, pol2-M644G* strains were identified and tested for a growth defect by spot assays in the presence or absence of the auxin hormone indole-3-acetic acid (IAA) (Figure 4.17). In the absence of IAA, the *pol2-M644G, RNH202-AID-9MYC, pADH1-Os.TIR1* strains did not exhibit a growth defect when compared to a wild-type control or the parental strains. When exposed to IAA, there was a negligible growth defect in comparison to the control strains. This was less severe than anticipated based on previous results in which RNase H2 activity had been lost either through deletion or tagging of the *RNH201* gene. Based on this result, it was postulated that RNase H2 activity was not hindered by the AID tag on its H2B subunit in the absence of IAA. However, further tests were not performed to assess the level of ribonucleotide incorporation in these strains due to the pursuit of alternative strategies. If further work could show an IAA-dependent increase in ribonucleotide levels in the genome, these strains could potentially be used for the polymerase-usage-based approach.

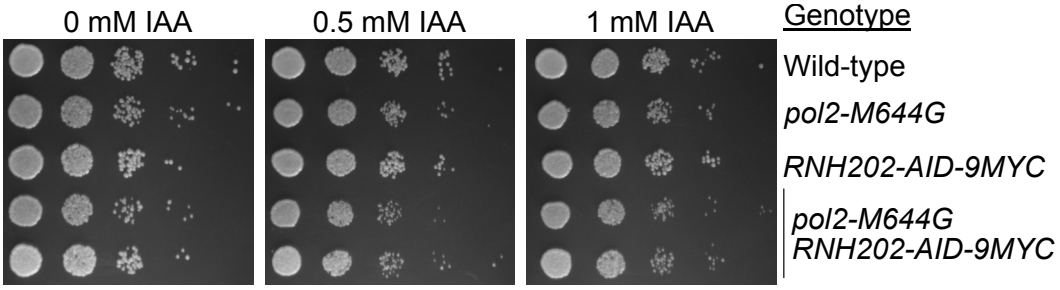


Figure 4.17*: Spot assays to assess growth of* pol2-M644G, RNH202-AID-9MYC, pADH1-Os.TIR1 *strains in the presence and absence of IAA. Ten-fold serial dilutions of strains with the indicated genotypes were grown on medium supplemented with the specified concentrations of IAA (all strains with* RNH202-AID-9MYC *also had* pADH1-Os.TIR1*). Plates were imaged after one day of incubation at 30 °C. Two isolates with the* pol2-M644G, RNH202-AID-9MYC, pADH1-Os.TIR1 *genotype were tested. These two strains showed similar growth to the wild-type or parental strains in the absence of IAA and a very slight defect in the presence of 0.5-1 mM IAA.*

4.5.3 Generating strains with inducible pol2-M644G

The possibility of controlling one of the mutant polymerases as an alternative to regulating RNase H2 activity to conditionally increase genomic ribonucleotide levels was considered. The aim was to generate a system in which a mutant replicative polymerase could be conditionally expressed such that it would compete with the endogenous wild-type polymerase during a single round of genome duplication. This required the integration of a mutant polymerase expression cassette at a distinct locus in the genome. First, an appropriate expression cassette had to be designed and constructed. To achieve this, the Type II restriction enzyme-based "hierarchical assembly system" compiled and described by (Lee et al. 2015) was used in a multi-step process (Figure 4.18). This required the amplification of the *pol2-M644G* gene, while mutating any *Bsm*BI, *Bsa*I, and *Not*I enzyme recognition sites to avoid inappropriate digestion during the assembly stages. The *pol2-M644G* gene in the W303 *S. cerevisiae* background has two *Bsa*I sites and two *Bsm*BI sites. Therefore, primers were designed to amplify the sequence in five segments to remove these sites (Figure 4.18 A). Once all fragments had been amplified and sequence verified, they were assembled into a universal entry vector as described in (Lee et al. 2015) (Figure 4.18 B). This meant that the *pol2-M644G* gene could then be combined with any other parts from the toolkit, including a range of different promotors and terminators. For the conditional expression cassette, the following assembly was produced: *pGAL1-pol2-M644G-3xFLAG-6xHis-tTDH1* in a plasmid designed for integration at the *ura3* locus (Figure 4.18 C). *Not*I digestion was used to linearise the cassette plasmid prior to transformation of wild-type and *rnh201Δ* strains, thus allowing chromosomal integration by homologous recombination.

Once transformants had been identified by selection on medium lacking uracil, a Western blot was used to verify Pol2-M644G-3xFLAG-6xHis protein expression in cultures grown in galactose (Figure 4.19). As a control, each strain was also grown in raffinose to verify that the protein was not expressed in non-inducing conditions. An anti-FLAG antibody was used to detect FLAG-tagged protein in the samples. No FLAG-tagged proteins were detected in the raffinose samples, but there was a clear band greater than 180 kDa in both samples when

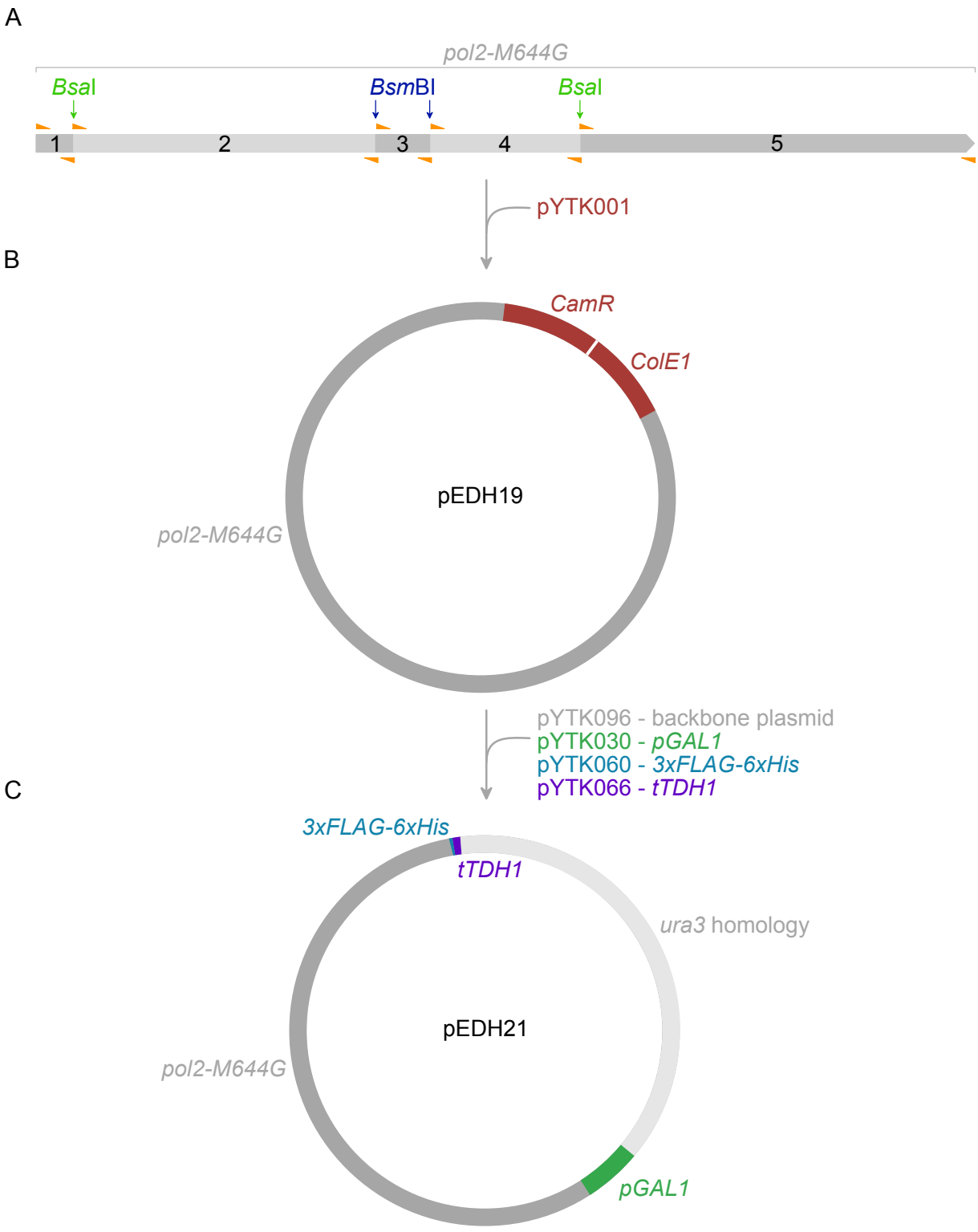


***Figure 4.18**: Schematic of the hierarchical assembly of* pGAL1-pol2-M644G-3xFLAG-6xHis-tTDH1 *in a* ura3 *integration cassette. (**A**) First, the* pol2-M644G *DNA was amplified as five fragments (labelled 1-5) using primers in the positions indicated by the orange arrows. These primers were designed to incorporate silent mutations to remove the restriction enzyme sites shown. (**B**) After repeated* BsmBI *digestion and ligation cycles, the five fragments were cloned into a part entry vector (pYTK001), producing the* pol2-M644G *part plasmid (pEDH19). This*

*encoded a chloramphenicol resistance marker (*CamR*) and the* ColE1 *origin of replication for expression in* E. coli, *in addition to the complete* pol2-M644G *sequence. (**C**) pEDH19 was then combined with a backbone vector (pYTK096) that included homology to* ura3 *to enable integration at this locus in the genome, and vectors containing the* pGAL1 *promoter (pYTK030), a* 3xFLAG-6xHis *tag (pYTK060) and a* tTDH1 *terminator (pYTK066). The parts were assembled in a repeated* BsaI *digestion and ligation cycle to form the illustrated plasmid (pEDH21).*

grown in galactose. The anticipated size of Pol2-M644G-3xFLAG-6xHis was approximately 260 kDa, so the Western blot result was consistent with galactose-dependent expression of Pol2-M644G-3xFLAG-6xHis in these strains.

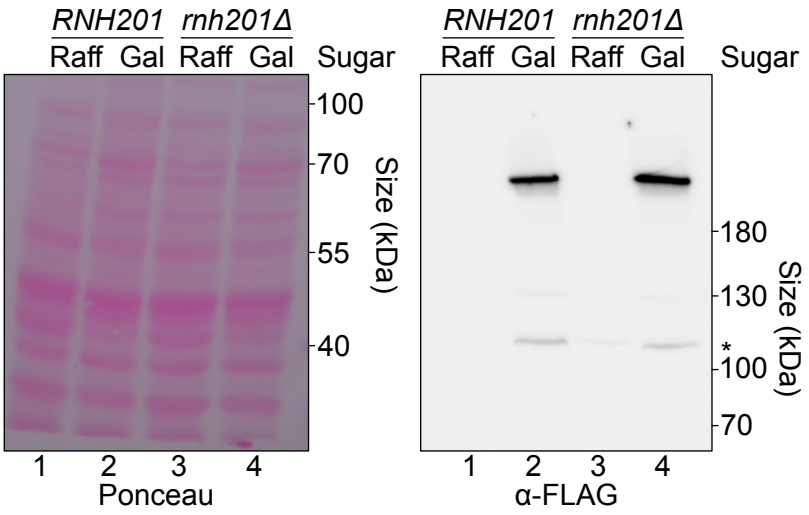


Figure 4.19*: A Western blot to detect FLAG-tagged Pol2-M644G. Strains with* pGAL1-pol2-M644G-3xFLAG-6xHis-tTDH1 *and the specified* RNH201 *genotype were grown with either raffinose (Raff) or galactose (Gal) as the sugar source. Total protein extract was size separated in a polyacrylamide gel and transferred to a membrane by Western blot. The Ponceau-stained membrane is shown as the loading control. When probed with an anti-FLAG antibody, signal was detected in bands >180 kDa (uppermost marker) in the samples grown in galactose but not in raffinose. * indicates FLAG-tagged protein degradation products.*

As the *pGAL1-pol2-M644G-3xFLAG-6xHis-tTDH1* strains were shown to express mutant polymerase ε in the presence of galactose, a spot assay was performed to determine whether the strains showed a growth defect in repressive or inducing conditions (Figure 4.20). The parental strains used for the transformations (both with endogenous wild-type *POL2* and either *RNH201* or *rnh201Δ*) were used as controls. Under both repressive (dextrose) and inducing (galactose) conditions, the *pGAL1-pol2-M644G-3xFLAG-6xHis-tTDH1* strains showed similar levels of growth to the parental strains. This suggested that the *pGAL1-pol2-M644G-3xFLAG-6xHis-tTDH1* genotype and the expression of the mutant polymerase did not confer a growth defect in either the presence or absence of the RNase H2 catalytic subunit. Thus, initial tests suggested that these strains may be suitable for the polymerase-usage-based approach. However, further characterisation would be necessary, including determination of the relative levels and activities of the wild-type and mutant Pol2 protein, and the extent of ribonucleotide incorporation during leading strand synthesis.

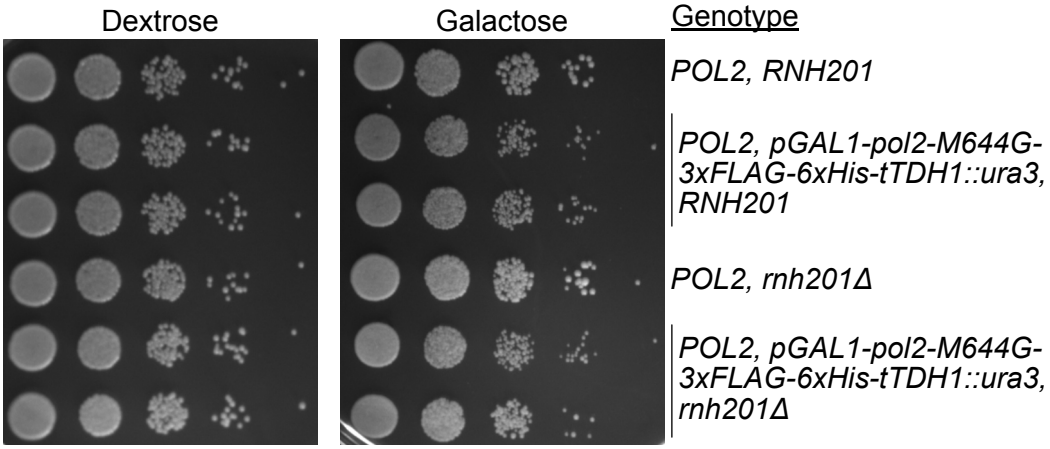


Figure 4.20*: Spot assays to assess growth of strains with* pGAL1-pol2-M644G-3xFLAG-6xHis-tTDH1 *in either repressive or inducing conditions. Two isolates of each* pGAL1-pol2-M644G-3xFLAG-6xHis-tTDH1 *strain (either with wild-type* RNH201 *or* rnh201Δ*) were grown in a ten-fold dilution series on medium containing either dextrose (represses expression from* GAL *promoters) or galactose (drives expression from* GAL *promoters). The parental wild-type and* rnh201Δ *strains were included as controls. Both plates were incubated at 30 °C and imaged after one day (dextrose) or two days (galactose). The* pGAL1-pol2-M644G-3xFLAG-6xHis-tTDH1 *strains showed comparable growth to the parental strains on both sugar sources.*

4.6 Summary

To summarise, a number of approaches for regulating RNase H2 or mutant polymerase levels were investigated in both *S. pombe* and *S. cerevisiae*. Attempts to conditionally deplete RNase H2 activity through a hormone binding domain in *S. pombe* did not result in a detectable increase in ribonucleotides in genomic DNA. Similarly, preliminary work testing transcriptional control using the *nmt81* promoter in *S. pombe* did not yield promising results.

It was shown that the presence of a 3xminiAID tag at the C-terminus of the RNase H2A subunit in *S. cerevisiae* led to an auxin-independent growth defect in *pol2-M644G* strains. An equivalent growth defect was not observed in strains with an AID tag at the C-terminus of the RNase H2B subunit, indicating that this strain could be further investigated for conditional increases in ribonucleotide levels in the genome. Finally, strains with an inducible *pol2-M644G* allele were produced and shown not to have a growth defect, again suggesting that these strains might be suitable for future work. However, due to promising contemporaneous work on the Okazaki fragment-based approach (Chapter 5), no further work was performed towards the polymerase-usage-based approach.

Chapter 5: Optical mapping of Okazaki fragments in single molecules

5.1 Introduction

In view of the challenges associated with producing strains suitable for optical mapping of polymerase usage discussed in Chapter 4, an alternative approach was explored for mapping replication dynamics in ultra-long single molecules. This approach also involved the labelling of nicks that mark replication. However, this method (Figure 5.1) was based on an Okazaki fragment (OF) sequencing methodology developed in budding yeast (Smith and Whitehouse 2012; Smith, Yadav, and Whitehouse 2015).

The 5' to 3' directionality of DNA synthesis results in asymmetric replication forks and semi-discontinuous duplication (Okazaki et al. 1968; Mitra and Kornberg 1966; Sakabe and Okazaki 1966; Snedeker, Wooten, and Chen 2017). The leading strand is synthesised continuously, while the discontinuous synthesis of the lagging strand as short OFs creates the need for additional maturation steps to complete nascent lagging strand synthesis (Balakrishnan and Bambara 2013). This process includes strand displacement of the RNA-DNA primer of the preceding OF by the lagging strand polymerase, Pol δ (Garg et al. 2004; Maga et al. 2001). The resulting flap is cleaved by nucleases (Garg et al. 2004; Maga et al. 2001; Jin et al. 2003; Kao et al. 2004), producing in a nick in the DNA backbone at the junction between two adjacent OFs. In *S. cerevisiae*, these nicks are sealed by DNA ligase I (encoded by *CDC9*) (Johnston and Nasmyth 1978). Therefore, it was hypothesised that removal of Cdc9 activity for a single S phase would result in daughter molecules with nicks at each OF junction in the nascent lagging strand. Such nicks could act as substrates for nick translation reactions to label the lagging strand with a fluorescently-labelled nucleotide for optical mapping.

The same experimental requirement of regulatable enzyme activity was also a feature of the OF approach, not only to limit nicking to just the nascent strand, but also because DNA ligase

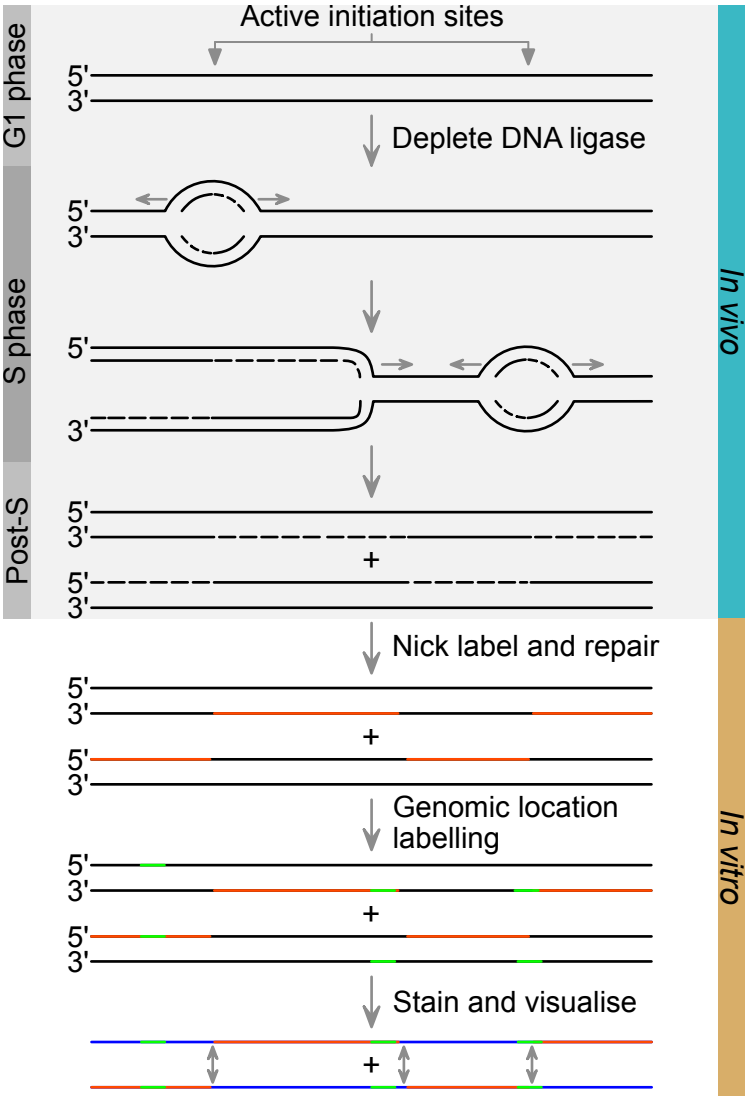


***Figure 5.1**: Outline of the experimental approach to obtain OF information in ultra-long single molecules by optical mapping. This requires a strain in which DNA ligase activity can be regulated. Cells are arrested in G1 phase and the ligase is depleted before release into S phase. After a single round of replication in the absence of DNA ligase activity, DNA is isolated and the unligated OFs are labelled with a fluorescent nucleotide (orange), and the translated nicks are then repaired. After labelling and ligating the nicks, the molecules can then undergo genomic location labelling (green) by treatment with DLE-1. Finally, the molecules are stained with YOYO-1 (blue) and can be visualised by Saphyr. The grey double-headed arrows mark the predicted transitions that might occur in the two theoretical molecules, from regions of OF label to regions of no OF label. Such transitions could mark sites of replication initiation or termination.*

I is essential for viability in *S. cerevisiae* (Giaever et al. 2002) and so cannot be deleted. This regulation has been achieved in several published studies, in which DNA ligase levels were conditionally depleted through the control of gene expression and/or protein degradation (McGuffee, Smith, and Whitehouse 2013; Smith and Whitehouse 2012; Kubota et al. 2013). Therefore, there was existing evidence that DNA ligase I levels could be controlled in *S. cerevisiae*. The results described in this chapter involved work with a published strain (Kubota et al. 2013) in which DNA ligase levels could be regulated by auxin-induced degradation.

5.2 Detecting DNA fragment size decreases when DNA ligase I is depleted in *S. cerevisiae*

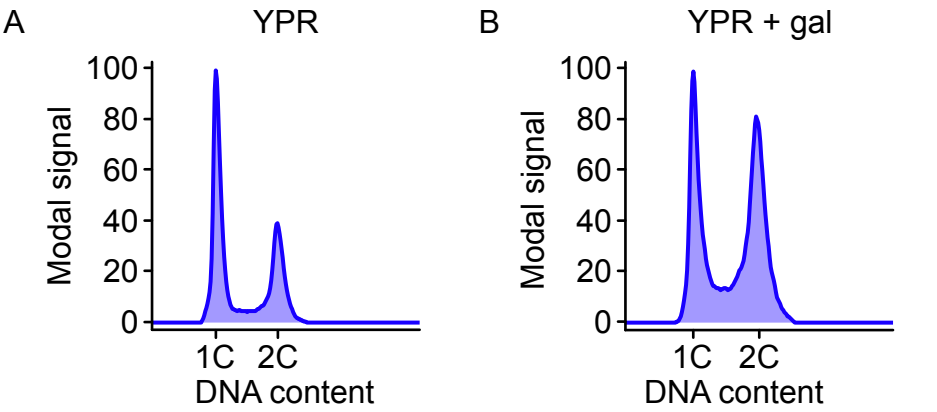
A *pGAL-Os.TIR1, CDC9-IAA17* (Cdc9-AID) strain was kindly provided by Dr. Kubota's group (University of Aberdeen), along with its wild-type parent. Growth of the Cdc9-AID strain in galactose-containing medium supplemented with the auxin indole-3-acetic acid (IAA) induces Tir1 expression to enable IAA-dependent degradation of Cdc9-AID. To observe the effect of Cdc9 degradation on cell growth, culture doubling times were measured in a range of conditions (Table 5.1). Dextrose is the preferred carbon source and represses expression from *GAL* promoters. The culture grown in dextrose had the shortest doubling time, as anticipated. The doubling time increased when grown in raffinose (non-inducing) and, to a slightly lesser extent, in galactose (inducing). Addition of IAA to cells cultured in raffinose did not notably impact the doubling time. However, IAA slowed cell growth in galactose. To assess growth over more generations under these conditions, each culture was diluted to reach an OD of 0.5 after 14.5 h of growth, based on the measured doubling times. Cultures with an OD > 0.5 after this time (non-induced, no IAA) suggested that the doubling time was slightly faster than that measured for the growth curve, while cultures with an OD < 0.5 (induced ± IAA, and non-induced + IAA) suggested slower growth. The slower growth of the culture grown in galactose without IAA could have resulted from the increased transcriptional and translational burden caused by Tir1 expression, and/or IAA-independent Cdc9 degradation. The greatest changes were observed in the cultures grown in IAA, with the Tir1-expressing cells showing very little growth over the 14.5 h. This indicated greatly perturbed growth upon Cdc9 depletion. The

***Table 5.1**: Doubling times of a* pGAL-Os.TIR1, CDC9-IAA17 *strain grown in different sugars ±IAA. Cells grew fastest in dextrose, which represses expression from* GAL *promoters and is the preferred carbon source. The doubling time increased when grown in raffinose and galactose compared to dextrose, with the growth in galactose slightly faster than in raffinose. Addition of IAA to non-induced cells did not increase the doubling time, but it did when Tir1 was induced. To assess growth over a longer period, cultures were diluted to be at an OD of 0.5 14.5 h later, based on the measured doubling times. The non-induced culture reached an OD slightly greater than expected, while the converse was observed for the induced culture. The presence of IAA in either non-induced or induced conditions considerably slowed growth, with the greatest effect in inducing conditions.*

Sugar	Dextrose	Raffinose	Galactose (+ raffinose)	Raffinose	Galactose (+ raffinose)
Tir1 expression	Repressed	Non-induced	Induced	Non-induced	Induced
+ 0.5 mM IAA?	No	No	No	Yes	Yes
Doubling time (min)	78	120	110	122	131
OD 14.5 h after dilution	Not measured	0.582	0.410	0.198	0.027

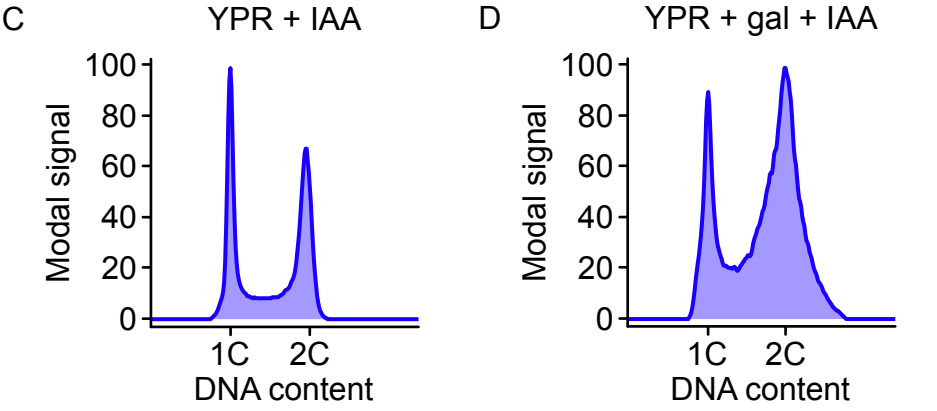
slower growth in non-inducing + IAA conditions was most likely a result of leaky Tir1 expression in raffinose causing Cdc9 degradation, and/or an independent detrimental effect of IAA on cells.

To determine whether cells were accumulating in one of the cell cycle stages as a result of Cdc9 degradation, samples were harvested from these cultures for analysis of DNA content by flow cytometry (Figure 5.2). This showed that the proportion of cells in G1 phase decreased when galactose was added to the culture. This was even more evident under conditions of Cdc9 degradation, with the majority of cells accumulating in late S phase and G2. These changes were consistent with hindered mitotic progression due to genome instability from the increased frequency of unresolved nicks.



Cell cycle phase	% of population
G1	29.2
S	16.6
G2	17.9

Cell cycle phase	% of population
G1	16.7
S	23.8
G2	26.3



Cell cycle phase	% of population
G1	22.7
S	27.7
G2	28.1

Cell cycle phase	% of population
G1	8.5
S	17.3
G2	23.3

***Figure 5.2**: Histograms of DNA content in asynchronous* pGAL-Os.TIR1, CDC9-IAA17 *cultures grown in different conditions. The DNA content of cells cultured overnight in the indicated conditions were analysed by flow cytometry to assess the population distribution in different cell cycle stages. The Cell Cycle tool in FlowJo was used to determine the proportion of each population in the indicated cell cycle phases. (**A**) Cells grown with 2% raffinose as the carbon source (YPR for yeast extract-peptone-raffinose medium). (**B**) Cells grown in YPR supplemented with 2% galactose (gal). (**C**) Cells grown in YPR supplemented with 0.5 mM IAA. (**D**) Cells grown in YPR supplemented with both 2% galactose and 0.5 mM IAA.*

To ascertain whether growth in galactose and IAA did deplete Cdc9, thereby reducing OF ligation, DNA fragment size distributions were analysed to detect an anticipated increased abundance of OF-sized molecules. Cells were arrested in G1 phase before induction of *TIR1* expression and IAA addition. Cells were incubated with IAA before release into S phase to enable Cdc9 degradation before the onset of genome replication. Flow cytometry was used to assess cell cycle arrest and subsequent S phase progression (Figure 5.3). This showed that cells were arrested in G1 phase and that the arrest was maintained during the galactose and IAA additions. The bulk population's DNA content began to increase 20-30 mins following

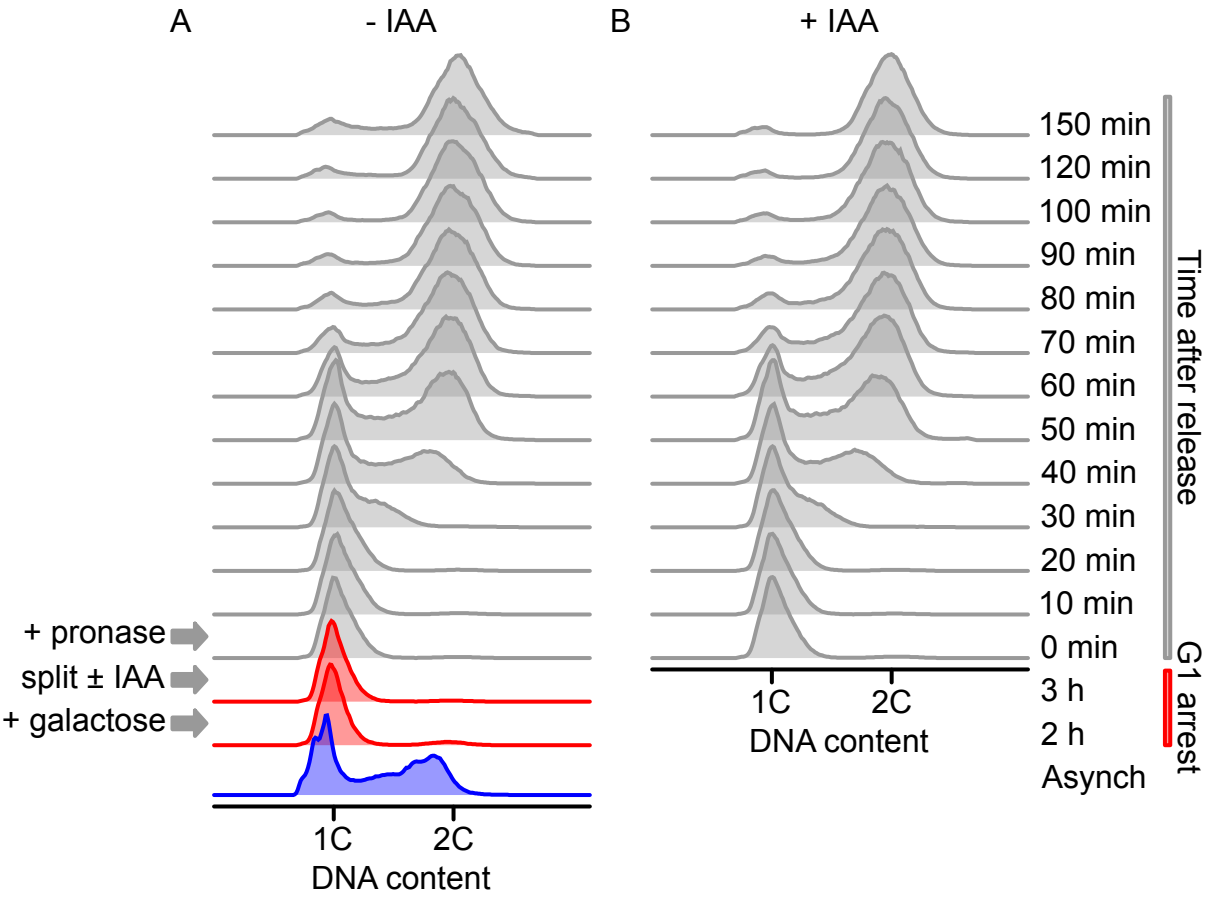


***Figure 5.3**: Histograms of DNA content in a* pGAL-Os.TIR1, CDC9-IAA17 *cell cycle arrest and release experiment in the presence or absence of IAA. (**A**) An asynchronous culture was arrested in G1 phase for 2 h before addition of galactose. Following 1 hr of G1 arrest in galactose, the culture was split in two, with one half receiving IAA (**B**) and the other not (**A**). Cultures were incubated for 1 hr before release from the G1 arrest. DNA content was used to assess the total DNA content of the populations at the indicated time points.*

release from the G1 arrest, consistent with entry into S phase. The majority of cells appeared to have reached a DNA content of 2C by 90 mins after release from the G1 block.

To determine whether there was a detectable increase in the abundance of OF-sized fragments following genome duplication in Cdc9-depleted cells, DNA was extracted from cells harvested at 0 and 90 min after release from the G1 arrest and size separated by denaturing gel electrophoresis (Figure 5.4 A). Two sets of samples were used to estimate the background levels of nicking – first, DNA from the culture that had not been treated with IAA (Figure 5.3 A), and second, DNA from cells at the point of release from G1, when DNA replication had not begun. All samples showed equivalent levels of nicking with the exception of DNA from cells treated with IAA and harvested 90 min following release from G1. This sample showed an increase in signal from molecules less than 1 kb in length compared to the other samples (Figure 5.4 B). This was consistent with a reduction in OF ligation in conditions where Cdc9 was expected to be depleted. Therefore, this strain met the requirement for a biological system in which replication-dependent nicks could be induced for a single S phase.

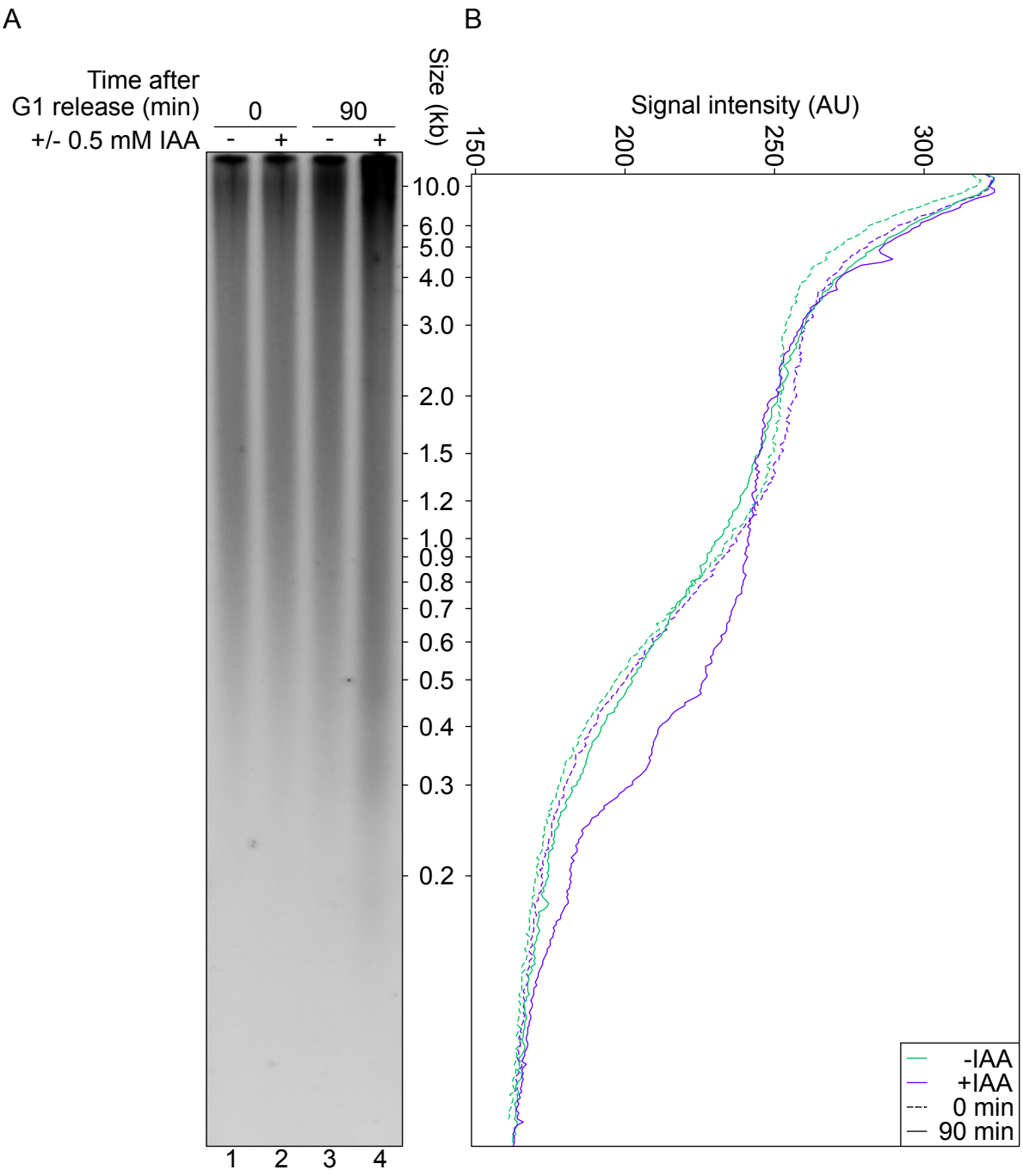


***Figure 5.4**: Denaturing gel electrophoresis of DNA from pre- and post-G1-release* pGAL-Os.TIR1, CDC9-IAA17 *cultures ± IAA. (**A**) DNA was purified from cells harvested 0 and 90 min after release from G1 phase (Figure 5.3), size separated on a denaturing gel and stained with Sybr Gold. Both 0 min samples and the 90 min -IAA sample showed very similar fragment size distributions. DNA from cells treated with IAA and harvested 90 min after release from G1 phase showed an increase in molecules shorter than 1 kb compared to the other samples. (**B**) Quantification of DNA signal intensity in a densitometry plot. An increase in signal intensity between 0.2-1.0 kb was observed in the 90 min +IAA sample, compared to the other samples.*

5.3 *In vitro* DNA labelling trials

For optical mapping, DNA labels such as incorporated fluorescently-tagged nucleotides are required. To characterise *in vitro* incorporation of Cy3-labelled dUTP into DNA, trial PCRs were performed. Aminoallyl-dUTP-Cy3 (henceforth referred to as dUTP-Cy3) was added to reactions in varying proportions to dTTP to establish whether *Taq* polymerase was able to use this labelled nucleotide as a substrate and whether this could be detected. The PCR products were run on a gel and visualised prior to any DNA staining, using Cy3-specific excitation and absorption wavelengths (Figure 5.5 A). Bands of increasing signal intensity appeared as the proportion of dUTP-Cy3 in the reaction increased, with no signal in the 0% dUTP-Cy3 sample, as anticipated. To detect total DNA, the gel was then stained with an intercalating dye and re-imaged (Figure 5.5 B). This showed a single band in all four samples in which polymerase had been added to the reaction, including the 0% dUTP-Cy3 sample. The bands matched the expectation of a 425 bp product. As the amount of dUTP-Cy3 in each reaction increased, the stained band intensity decreased. This suggested that increasing dUTP-Cy3 incorporation hindered amplification, thereby reducing the total amount of product.

Next, the ability to incorporate dUTP-Cy3 in a nick-translation reaction was tested to more closely resemble the genomic DNA labelling reaction. PCR was used to amplify a sequence with a single Nt.*Bsp*QI nick site (Figure 5.6 A). The product was treated with Nt.*Bsp*QI, followed by *Taq* polymerase with dNTPs including varying proportions of dUTP-Cy3:dTTP. The resulting molecules were run on a denaturing gel to separate the labelled strand from its complement and the Cy3 signal was visualised (Figure 5.6 B). Nick-translation in the presence of 20, 35 or 50% dUTP-Cy3 gave bands at approximately 800 bases, consistent with the anticipated length of 783 bases. Nick labelling with 100% dUTP-Cy3 (0% dTTP) also gave a Cy3-containing product which was smaller in size than those with lower proportions of dUTP-Cy3 (less than 700 bases). This suggested that the polymerase was unable to extend the full 136 bases in this reaction. The gel was then treated with ethidium bromide to enable visualisation of all DNA (Figure 5.6 C). This revealed that the majority of strands in each sample were just under 700 bases with an additional, less prevalent molecule of just under 600 bases

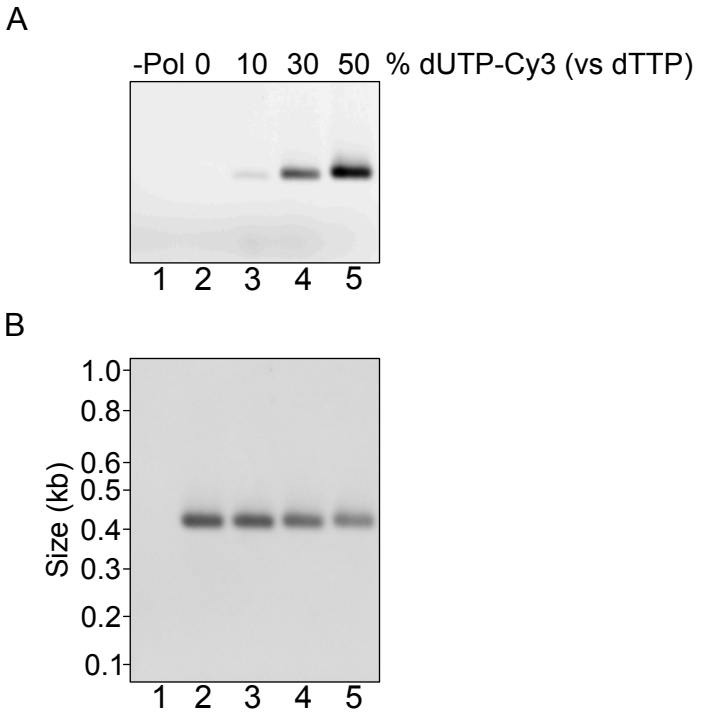


***Figure 5.5**: Detection of dUTP-Cy3 incorporated during PCR. (**A**) Visualisation of Cy3 signal prior to treatment with a DNA dye. Five PCRs were performed using the same template and starting reaction mix with varying proportions of dUTP-Cy3:dTTP, then run on a gel. The -Pol reaction had 30% dUTP-Cy3 and 70% dTTP but no* Taq *polymerase. The 0% dUTP-Cy3 reaction had 100% dTTP (the same concentration as each of the other dNTPs). Increasing the proportion of dUTP-Cy3 to dTTP resulted in stronger Cy3 signal in the detected bands. (**B**) The gel shown in (A) was stained with ethidium bromide and re-visualised to detect all DNA. Each of the bands was just over 400 bp in size, with the anticipated PCR product length being 425 bp. The band intensity decreased as the proportion of dUTP-Cy3 increased.*

in the nicked only and 100% dUTP-Cy3 samples. These apparent molecule lengths were inconsistent with the expectation based on the PCR amplicon length and the position of the nick site. This disparity could have been caused by partial nicking of the PCR product and/or inaccurate estimation of molecule sizes due to perturbed size separation of the molecular weight marker in a denaturing gel. However, the cause was not investigated. Nevertheless, it was concluded that it was possible to label a nick in DNA with varying proportions of dUTP-Cy3 in the reaction mix but that labelling with 100% dUTP-Cy3 inhibited extended elongation from the nick.

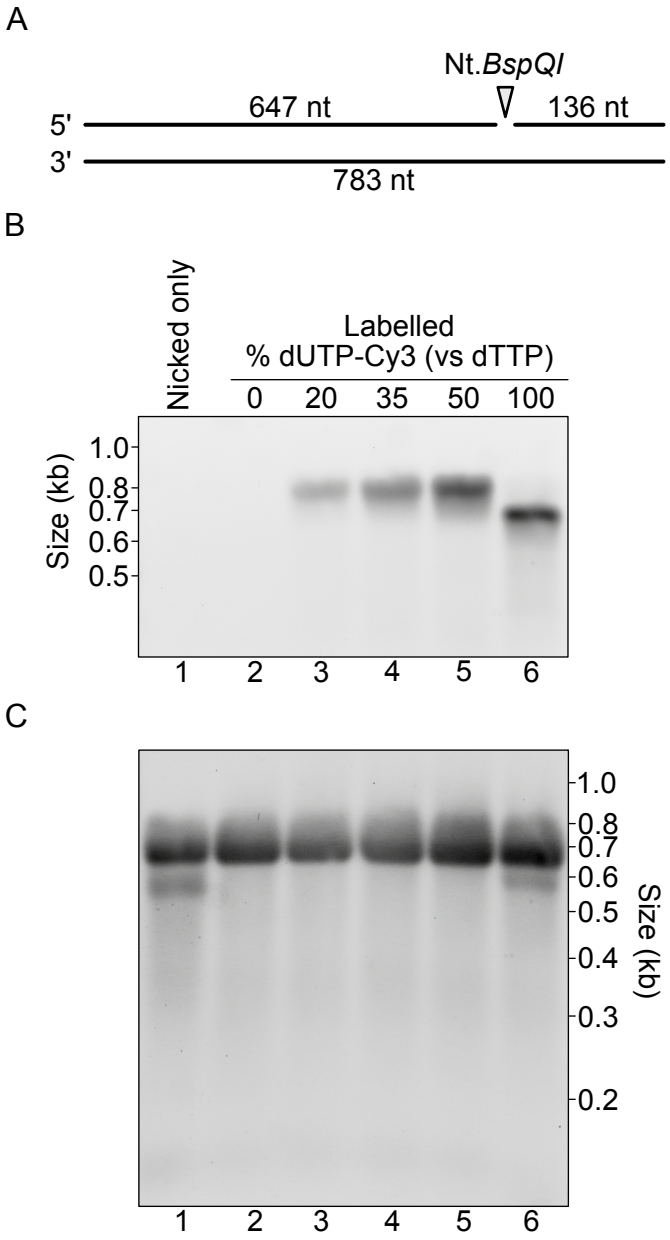


***Figure 5.6**: dUTP-Cy3 incorporation during nick-translation of a PCR product. (**A**) Schematic illustrating the location of the Nt.*BspQI *nick site in a 783 bp PCR product. (**B**) The PCR product in (A) was nicked with Nt.*BspQI *and nick translation reactions were performed to extend from this site, with varying proportions of dUTP-Cy3:dTTP. The samples were run on a denaturing gel and Cy3-labelled products were visualised prior to DNA staining. Reactions with 20-50% dUTP-Cy3 yielded products of approximately 800 bases in length, while the reaction with 100% dUTP-Cy3 gave a shorter product. (**C**) The gel in (B) was treated with ethidium bromide to stain all DNA. All samples showed strong bands just under 700 bases in length, with additional minor bands at just under 600 bases in the nicked only and 100% dUTP-Cy3 lanes (1 and 6).*

5.4 Labelling unligated Okazaki fragments *in vitro*

To determine whether unligated OF nicks could be labelled with dUTP-Cy3, a cell cycle arrest and release experiment was performed to acquire cells that had gone through a single S phase with Cdc9 depleted. A wild-type strain treated identically to the *pGAL-Os.TIR1, CDC9-IAA17* strain was used to provide an indication of background levels of nicking. Flow cytometry analysis of samples taken throughout the duration of this cell cycle experiment showed that the majority of cells were successfully arrested in G1 phase prior to the addition of galactose and IAA (Figure 5.7). Once the cultures were released from the G1 arrest, cells progressed through S phase with the majority of the population reaching a DNA content of 2C by 90 min after the release in both samples. Flow cytometry analysis of the samples harvested from 110 min to 180 min after release indicated that a proportion of cells in the wild-type culture had re-entered G1 phase and were continuing with cell cycle progression (Figure 5.7 A). In contrast, the *pGAL-Os.TIR1, CDC9-IAA17* culture remained arrested with a DNA content of 2C (Figure 5.7 B). This was consistent with impeded mitotic progression due to the genome-wide abundance of unligated OFs.

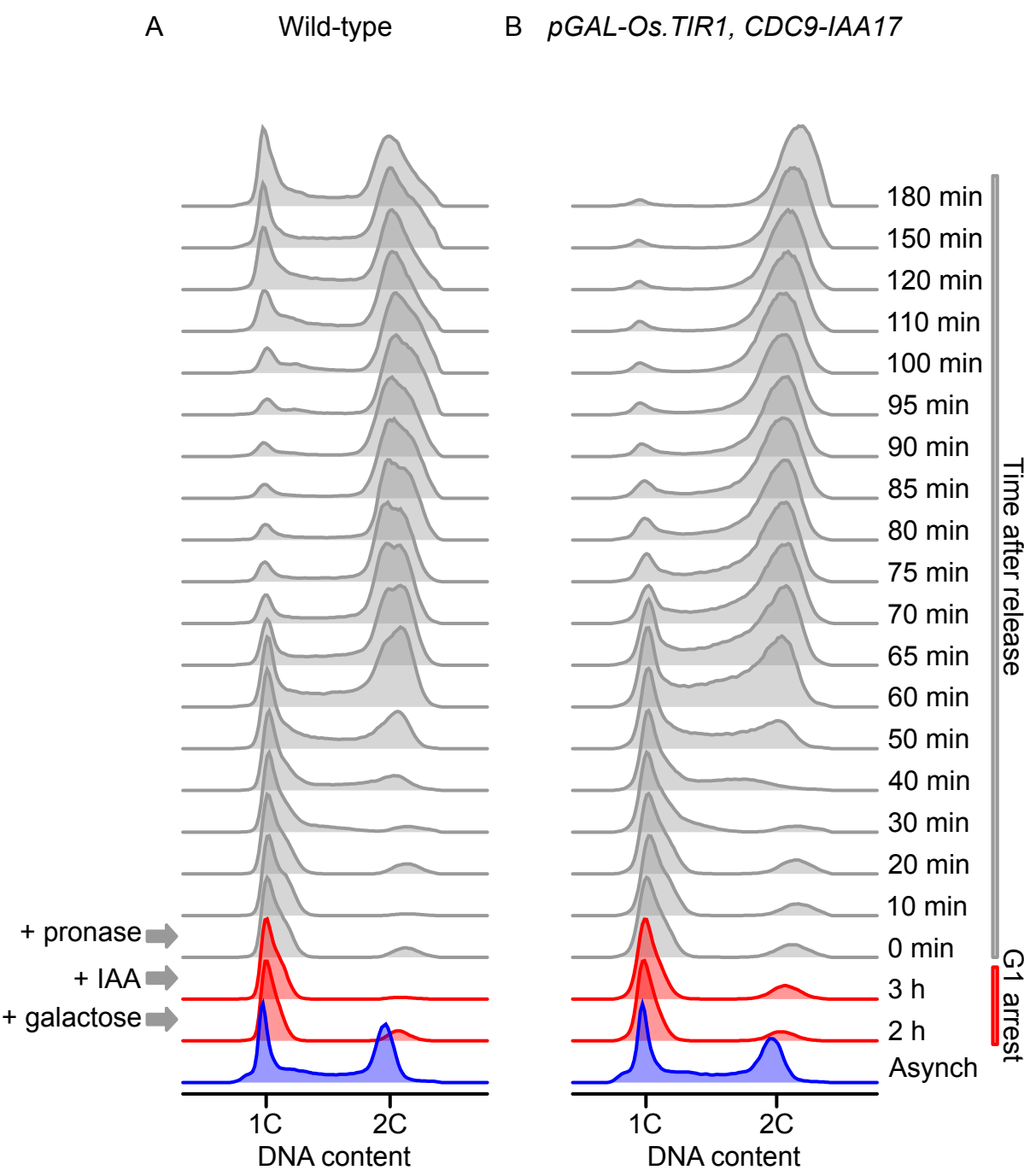


***Figure 5.7**: Histograms of DNA content in a wild-type and* pGAL-Os.TIR1, CDC9-IAA17 *cell cycle arrest and release. Both cultures were arrested in G1 phase with α-factor for 2 h before addition of galactose to induce* TIR1 *expression in the* pGAL-Os.TIR1, CDC9-IAA17 *strain. One hour after galactose addition, IAA was added to both cultures to trigger degradation of Cdc9-AID in the* pGAL-Os.TIR1, CDC9-IAA17 *strain. One hour after IAA addition, cells were released from the G1 arrest by addition of pronase and proceeded into S phase. Samples were harvested for DNA extraction from both cultures at 100 min after the release. (**A**) Flow cytometry analysis of samples from the wild-type culture. The majority of cells had a DNA*

*content of 2C by 75 min after the release and cells began to re-enter G1 phase by 110 min. (**B**) Flow cytometry analysis of samples from the* pGAL-Os.TIR1, CDC9-IAA17 *strain. The majority of cells in this culture had a DNA content of 2C by 90 min after release, but the population remained arrested at this stage in the cell cycle for the remainder of the experiment.*

DNA was extracted from cells harvested 100 min after release from the G1 arrest with the goal of labelling unligated OF nicks present in the *pGAL-Os.TIR1, CDC9-IAA17* genomic DNA. The DNA was treated with *Taq* DNA polymerase and dNTPs for labelling by nick translation. The dNTP concentrations (including the fluorophore-labelled dUTP) used in the BNG nick label reactions were unknown (proprietary), so an initial trial was performed using 25 μM of each dNTP. To label nicks for detection, dTTP was substituted with dUTP-Cy3 in a ratio of 65:35 in the reaction. The labelled genomic DNA was run on a denaturing gel and visualised to detect Cy3-labelled DNA (Figure 5.8 A). Cy3 signal was detected in the *pGAL-Os.TIR1, CDC9-IAA17* sample, but not that of the wild-type, consistent with specific labelling of unligated OFs. The molecule size distribution was longer than expected for individual OFs (approximately 165 bp) (Smith and Whitehouse 2012). This increased length of the Cy3-labelled fragments might have been due to incomplete Cdc9 depletion during S phase enabling some OF ligation. Alternatively, or additionally, the conditions used for nick translation might have allowed the fragments to be extended for long distances, thereby increasing the fragment lengths. Total DNA was then stained and visualised using Sybr Gold (Figure 5.8 B). This confirmed the presence of genomic DNA in the wild-type sample which was not detected by Cy3 labelling. The DNA from the *pGAL-Os.TIR1, CDC9-IAA17* strain showed greater fragmentation than that of the wild-type strain, consistent with the presence of unligated OFs due to Cdc9 depletion.

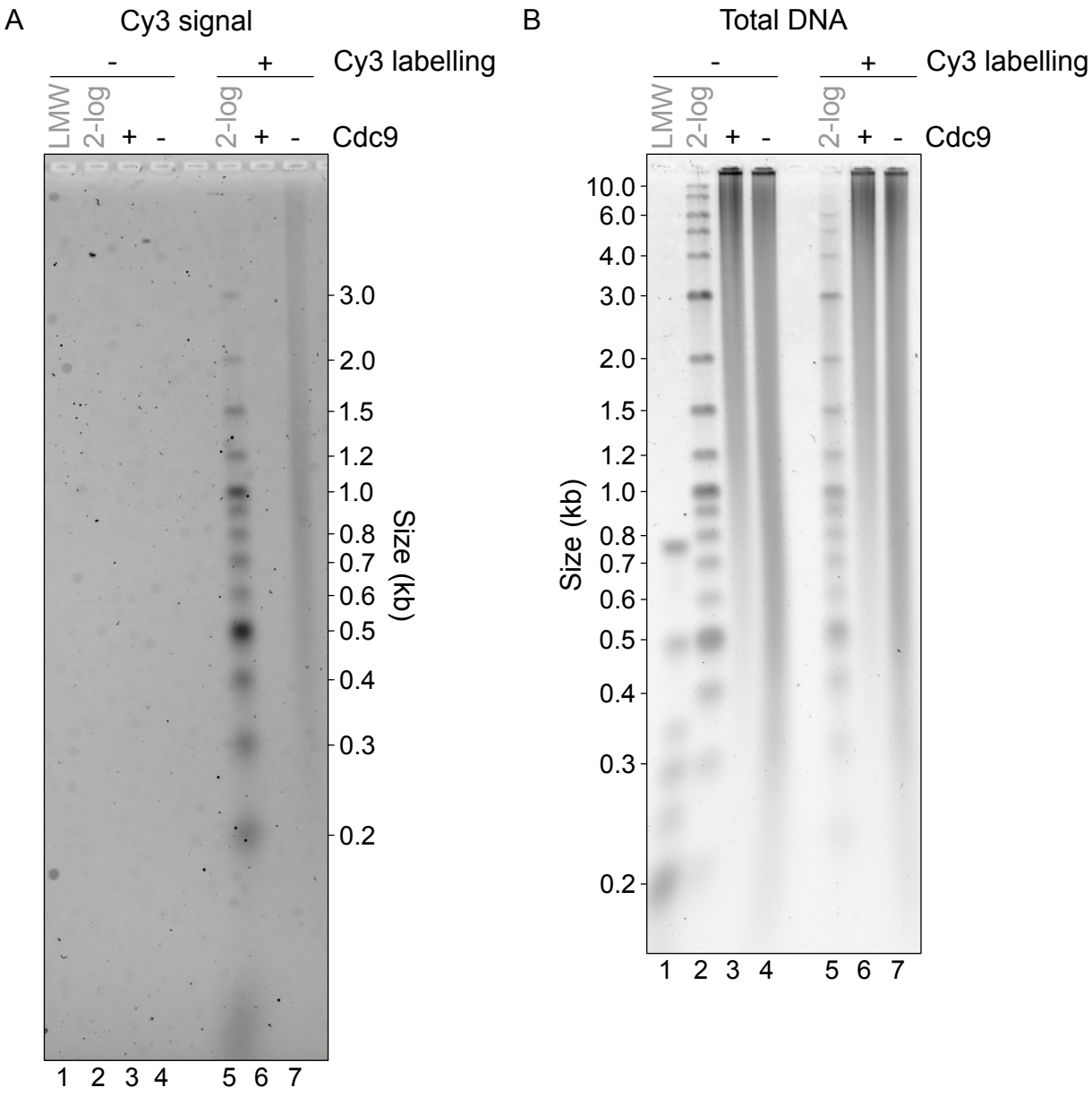


Figure 5.8*: Detection of dUTP-Cy3 incorporation during nick-translation of genomic DNA by gel electrophoresis. Genomic DNA was extracted from samples harvested 100 min after release from the G1 arrest (Figure 5.7) and size separated on a denaturing gel both before and after* in vitro *nick labelling with dUTP-Cy3. DNA from wild-type and* pGAL-Os.TIR1, CDC9-IAA17 *strains (labelled Cdc9 + and – to indicate normal and depleted Cdc9 levels, respectively) was analysed. DNA markers of known size (2-Log and LMW (Low Molecular Weight)) were run alongside the samples. (**A**) Visualisation of Cy3-labelled DNA showed signal in the end-labelled 2-Log ladder and the nick-translated* pGAL-Os.TIR1, CDC9-IAA17 *DNA, but not in the nick-translated wild-type DNA. The Cy3-labelled OFs ranged from approximately 400 bases to over 3 kb in size. (**B**) The gel shown in (A) was treated with Sybr Gold to stain total DNA. DNA*

was detected in both wild-type and pGAL-Os.TIR1, CDC9-IAA17 *samples. The* pGAL-Os.TIR1, CDC9-IAA17 *DNA showed a greater abundance of shorter fragments compared to the wild-type DNA.*

To identify conditions in which nicks labelled by dUTP-Cy3 incorporation could be repaired, *in vitro* treatment with *Taq* DNA ligase was tested. To obtain DNA samples, an arrest/release experiment was performed, similarly to that shown in Figure 5.7. Cells were harvested 90 min after the release from the G1 arrest, as this was previously the point at which the majority of cells in both samples had reached a DNA content of 2C. The DNA content of the samples was analysed by flow cytometry to confirm the G1 arrest and release (Figure 5.9). The sub-1C peak detected in the asynchronous samples, most prominently in the wild-type strain (Figure 5.9 A), was of unknown identity and was not detected in subsequent samples from this experiment. In both cultures, the majority of cells had reached a DNA content of 2C, consistent with completion of genome duplication, by 90 min after the release. However, a small 1C peak remained in both samples, corresponding to unreleased G1 cells. In addition, some cells had a DNA content between 1C and 2C in the *pGAL-Os.TIR1, CDC9-IAA17* culture, indicating that they had not yet completed genome replication (Figure 5.9 B).

DNA was purified from the samples harvested 90 min after release from the G1 arrest. One goal was to optimise the *Taq* DNA polymerase nick labelling conditions with an appropriate concentration of each dNTP. Conditions used in published studies had successfully labelled DNA for BNG analysis with 100% of the labelled nucleotide (and 0% dTTP) (Gabrieli et al. 2018b, 2017; Sharim et al. 2019; De Carli et al. 2016; McCaffrey et al. 2017). Therefore, this setup was used in subsequent experiments. The studies varied in the nucleotide concentrations, using either 100 nM or 600 nM of each dNTP. However, many of these studies were focussed on more localised regions of signal, rather than half of the nascent DNA, as expected here. Thus, the concentration was increased to ensure that each dNTP was available in excess with the aim of labelling every unligated OF. In this trial, 5 µM of each dNTP was added in the labelling reaction, with dUTP-Cy3 in place of dTTP. Following the labelling

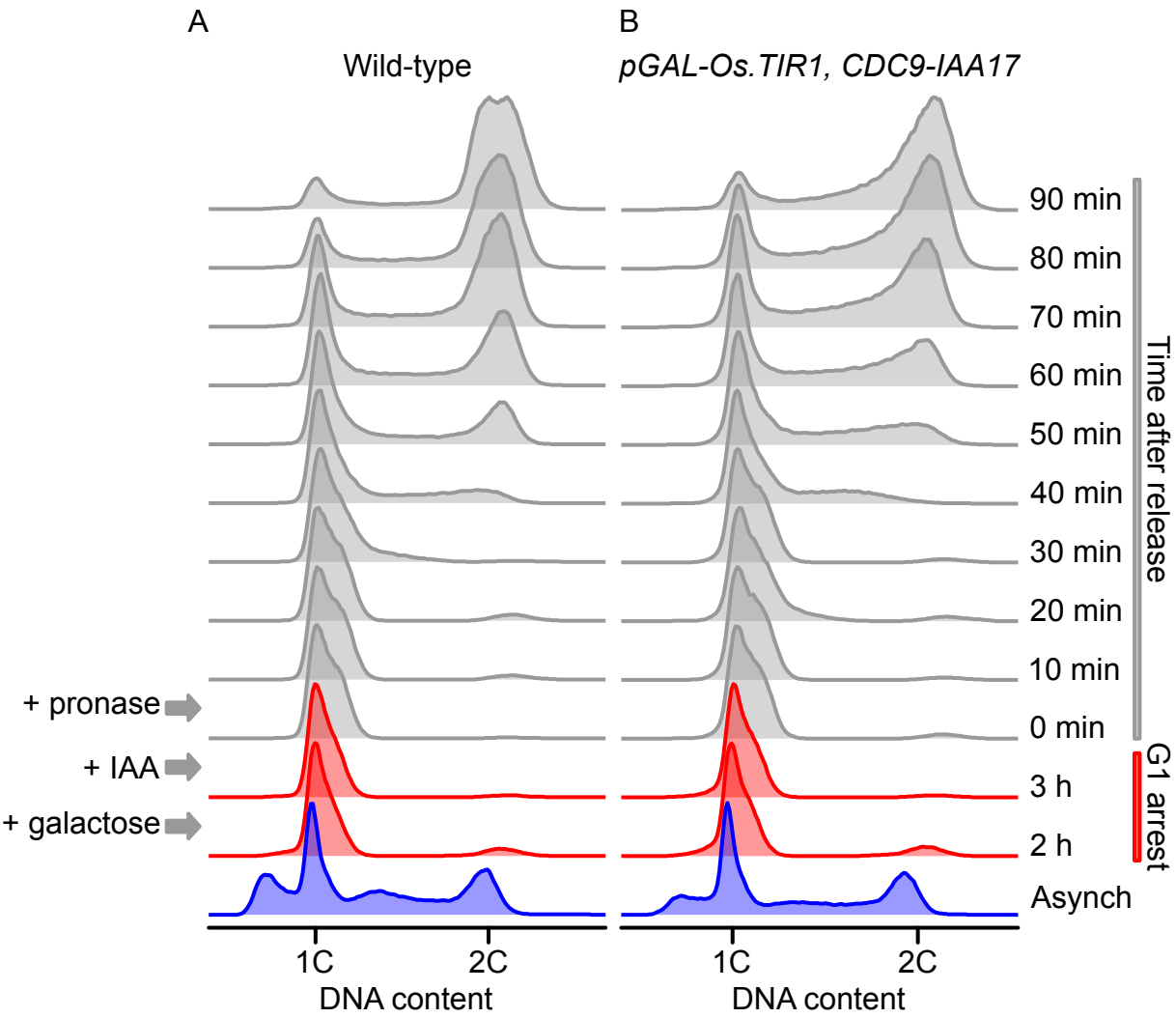


Figure 5.9*: Histograms of DNA content in a wild-type and* pGAL-Os.TIR1, CDC9-IAA17 *cell cycle arrest and release. As in Figure 5.7, with samples harvested for DNA extraction at 90 min after the release. (**A**) Analysis of samples from the wild-type culture. The majority of cells completed genome duplication by 90 min after the release. (**B**) Analysis of samples from the* pGAL-Os.TIR1, CDC9-IAA17 *strain. The majority of cells in this culture had replicated their genome by 90 min after release, although some cells had a DNA content between 1C and 2C.*

reaction, *in vitro* nick repair was attempted using *Taq* DNA ligase, which specifically seals nicks in DNA. It had previously been shown that restriction enzyme-induced nicks could be labelled and repaired, producing molecules suitable for optical mapping by Irys (section 4.2). Thus, as a positive control, wild-type DNA was nicked *in vitro* with the two nicking endonucleases Nt.*Bsp*QI and Nb.*Bbv*CI, and was subjected to the same labelling and repair reaction conditions as the *pGAL-Os.TIR1, CDC9-IAA17* DNA. The reaction products were run on a

denaturing gel and Cy3-labelled DNA was visualised (Figure 5.10 A). The restriction endonuclease nicking of wild-type DNA enabled detection of Cy3-labelled fragments that were generally longer than those detected in the *pGAL-Os.TIR1, CDC9-IAA17* DNA, consistent with the nick sites occurring much less frequently than unligated OF nicks. The three bands of signal observed within the smear were consistent with *Bsp*QI sites in the rDNA, of which there are three in each repeat unit (and no *Bbv*CI sites) in the sacCer3 reference. The *pGAL-Os.TIR1, CDC9-IAA17* DNA fragments in this experiment were mostly less than 1 kb in length, with a shorter fragment size distribution than that seen in Figure 5.8. This suggested that the labelling conditions used in Figure 5.8 did contribute to the longer fragments observed and that using 100% dUTP-Cy3 could be more optimal for achieving shorter tracts of label incorporation. Following treatment with *Taq* DNA ligase, the labelled wild-type DNA did show an upward shift in size distribution. The change in length distribution was clearer for the *pGAL-Os.TIR1, CDC9-IAA17* DNA, where the molecule size distribution shifted from the majority being under 1 kb, to most being longer than 3 kb (Figure 5.10 A). This was consistent with successful labelling and repair of unligated OF nicks under these conditions. Total DNA staining confirmed the presence of DNA in all lanes and the greater abundance of shorter fragments in DNA from the *pGAL-Os.TIR1, CDC9-IAA17* strain, compared to that from the wild-type strain (Figure 5.10 B).

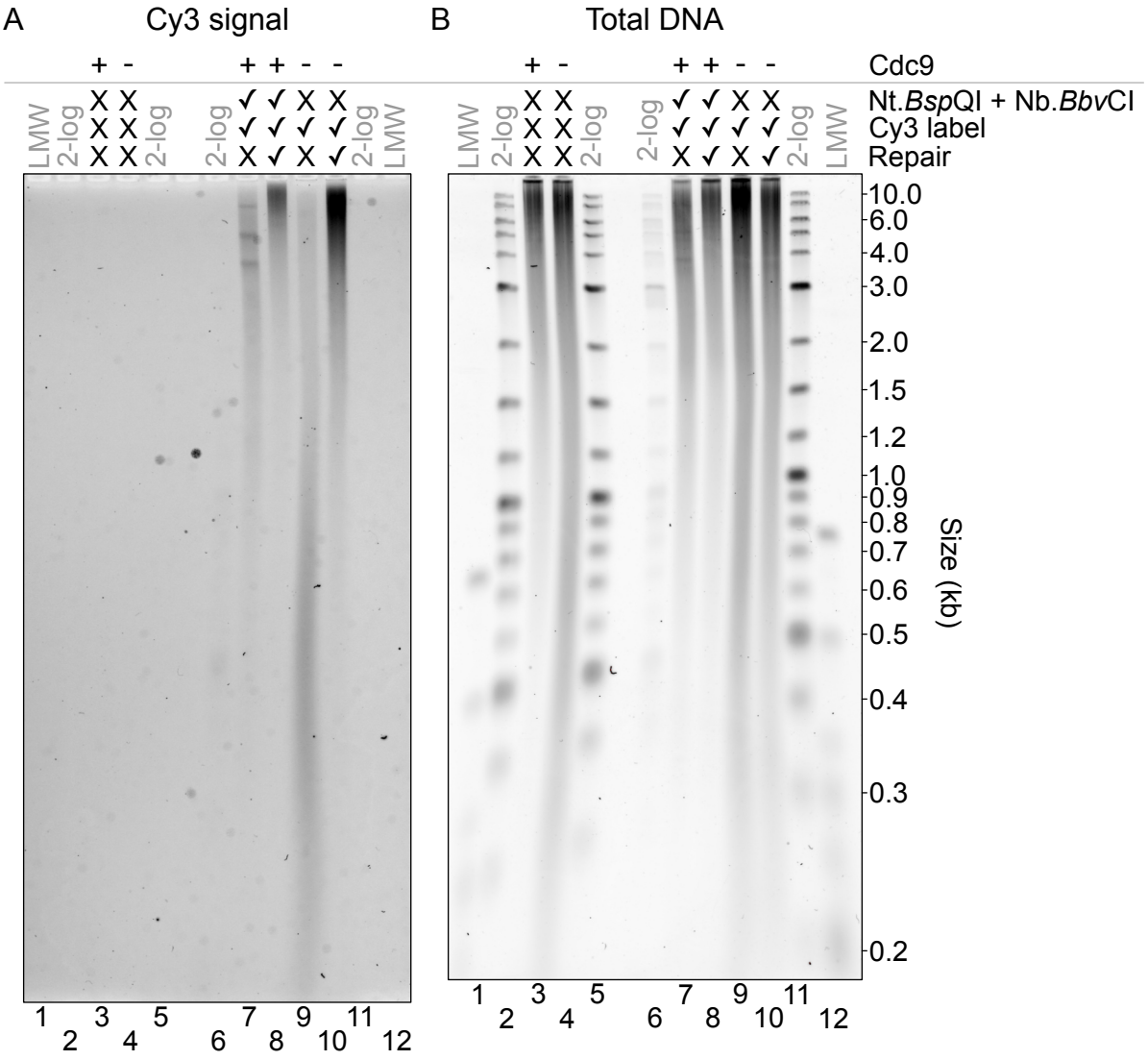


Figure 5.10*: Visualisation of dUTP-Cy3 incorporation during nick-translation of genomic DNA, followed by nick repair. DNA from a wild-type and the* pGAL-Os.TIR1, CDC9-IAA17 *strain was extracted from cells in which the majority of the population had a DNA content of 2C (Figure 5.9) and run on a denaturing gel before (lanes 3 and 4) or after (lanes 7-10)* in vitro *treatments. Cdc9 + and – indicate normal and depleted Cdc9 levels* in vivo*, respectively. The* in vitro *conditions (below the grey line) indicate whether DNA was nicked with Nt.*Bsp*QI and Nb.*Bbv*CI following extraction, labelled by nick translation with* Taq *DNA polymerase incorporation of dNTPs and dUTP-Cy3, and repaired by* Taq *DNA ligase. DNA markers of known size (2-Log and LMW) were run alongside the samples. (**A**) Visualisation of Cy3 signal. Untreated DNA did not show detectable Cy3 signal (lanes 3 and 4), while* in vitro *labelled DNA was detected (lanes 7-10). The wild-type DNA that was nicked and labelled showed slightly shorter fragments than the same DNA that was subsequently treated with DNA ligase. Similarly, the*

labelled pGAL-Os.TIR1, CDC9-IAA17 *DNA mainly consisted of fragments <1 kb in length, but treatment with DNA ligase shifted the fragment sizes to mostly >3 kb. (**B**) Visualisation of total DNA. Wild-type DNA had longer fragments than the* pGAL-Os.TIR1, CDC9-IAA17 *DNA, even after* in vitro *nicking (lanes 3, 7 and 8 compared to lanes 4, 9 and 10).*

All nick, label, repair trials described above had used genomic DNA extracted using the protocol described in (Smith, Yadav, and Whitehouse 2015). Given that optical mapping requires ultra-long molecules, subsequent trials involved suspending spheroplasts in agarose, forming plugs, to facilitate the isolation of high molecular weight molecules. The first goal was to assess the molecule size distribution following DNA extraction. The cell cycle arrest/release experiment was repeated (Figure 5.11), and cells harvested 90 min after the G1 release were spheroplasted and embedded in agarose for DNA isolation.

To assess the molecule size distribution of DNA isolated from the 90 min cells, DNA embedded in plugs, as well as DNA purified from plugs was analysed by pulsed field gel electrophoresis (PFGE, Figure 5.12). Banding patterns similar to the chromosomal markers in plugs from both the wild-type and *pGAL-Os.TIR1, CDC9-IAA17* strains indicated the presence of chromosomal-length DNA, with increased smearing observed in the plugs containing *pGAL-Os.TIR1, CDC9-IAA17* DNA, indicative of increased fragmentation upon Cdc9 depletion. Overall, this suggested that, while embedded in agarose, the chromosomes stayed largely intact during the cell lysis and DNA purification steps. However, once the DNA had been extracted from the agarose plugs, there was no longer a chromosomal banding pattern and the majority of the signal from both samples was found in a smear made up of molecules mostly shorter than 450 kb. This was consistent with DNA fragmentation during the purification from agarose. Despite this, the DNA was still of sufficiently high molecular weight to potentially be useful for optical mapping.

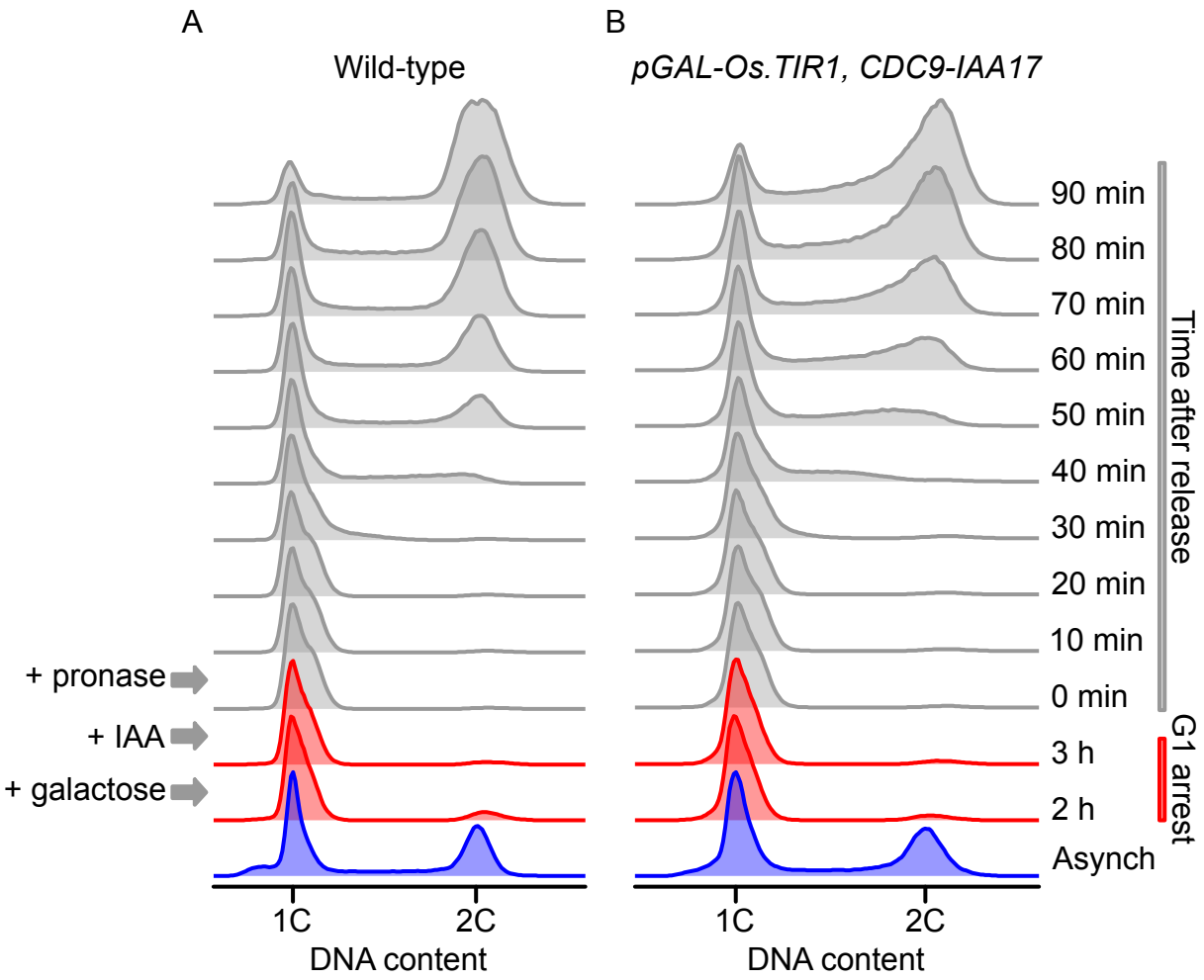


***Figure 5.11**: Histograms of DNA content in a wild-type and* pGAL-Os.TIR1, CDC9-IAA17 *cell cycle arrest and release, as in Figure 5.9. (**A**) Flow cytometry analysis of samples from the wild-type culture. The majority of cells had copied their genome by 90 min after the release. (**B**) Flow cytometry analysis of samples from the* pGAL-Os.TIR1, CDC9-IAA17 *strain. The majority of cell in this culture had a DNA content of 2C by 90 min after release, but some cells still had a DNA content between 1C and 2C.*

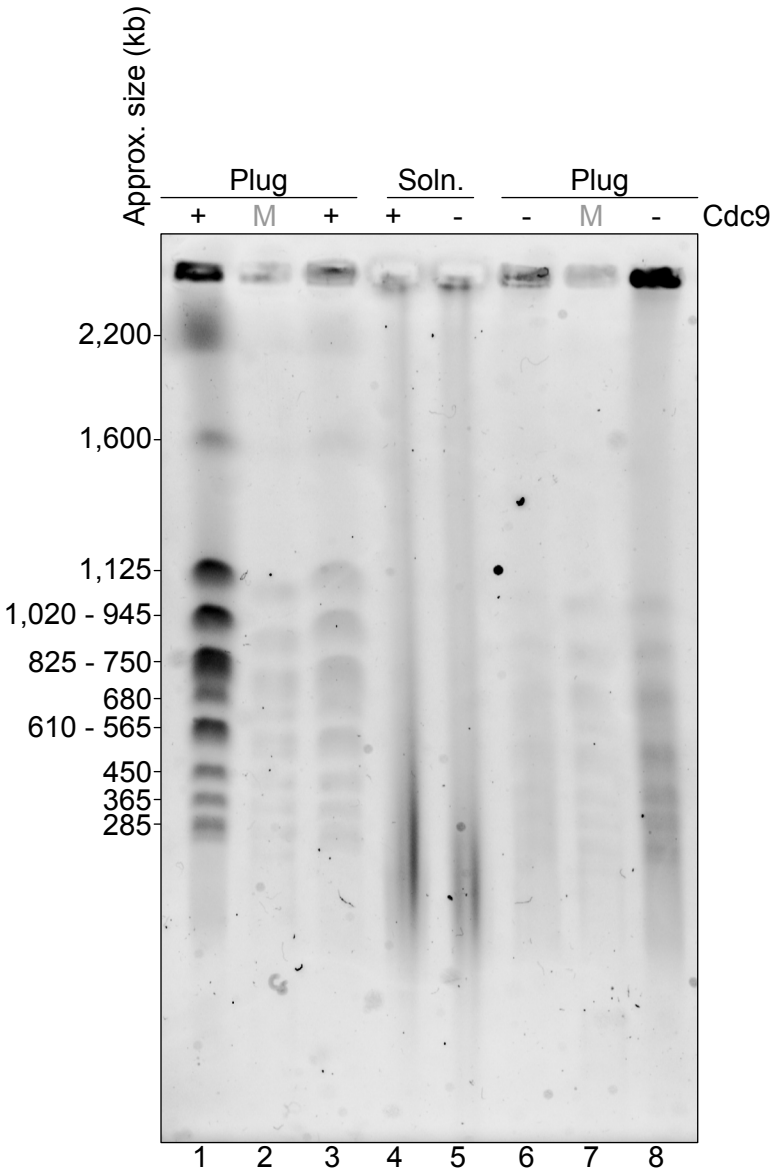


***Figure 5.12**: PFGE analysis of wild-type and* pGAL-Os.TIR1, CDC9-IAA17 *DNA. Cells harvested 90 min after release from a G1 arrest (Figure 5.11) were spheroplasted and embedded in agarose to form plugs. While embedded in the agarose plugs, cells were lysed, RNase A and proteinase K treated, and washed. DNA in the resulting plugs was size separated by PFGE (lanes 1, 3, 6, and 8) alongside commercial* S. cerevisiae *chromosomal markers (M) in plugs (lanes 2 and 7). Plugs from the wild-type strain ("Cdc9 +", lanes 1 and 3) showed chromosomal-length molecules, while plugs from the* pGAL-Os.TIR1, CDC9-IAA17 *strain ("Cdc9 -", lanes 6 and 8) showed chromosomal-length bands with increased smearing compared to the wild-type. Following extraction of both wild-type and* pGAL-Os.TIR1, CDC9-IAA17 *DNA from plugs (lanes 4 and 5, DNA in "soln."), the chromosomal-sized bands were*

lost and the DNA produced a smear of signal corresponding predominantly to molecules < 450 kb.

Next, it was necessary to establish whether the incorporation of fluorophore-labelled dUTP during nick translation from unligated OFs interfered with the genomic location DLE labelling. The cell cycle arrest/release experiment was repeated (Figure 5.13) to obtain DNA for labelling. Cells were harvested 90 min after release from the G1 arrest, when the majority of

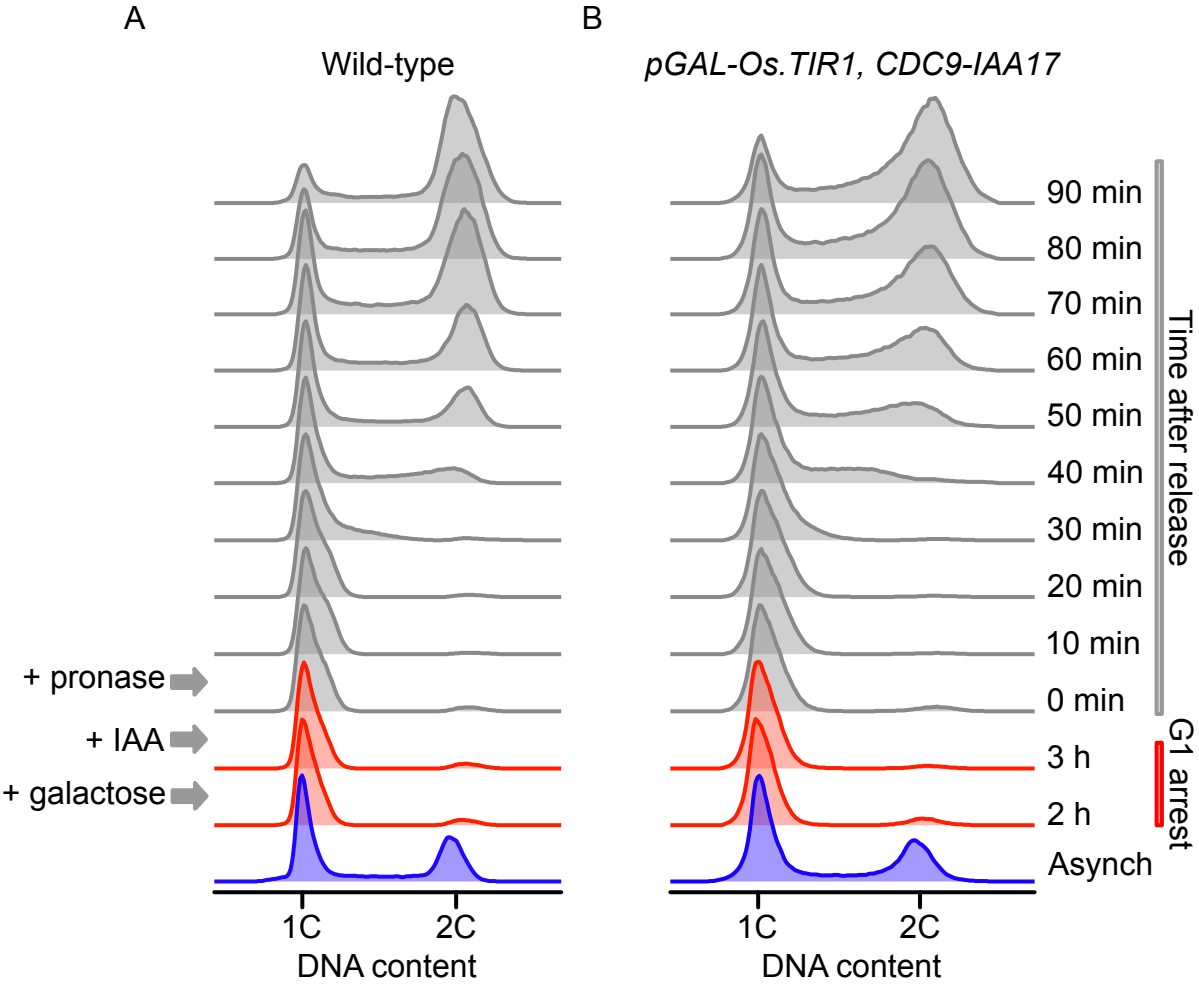


***Figure 5.13**: Histograms of DNA content in a wild-type and* pGAL-Os.TIR1, CDC9-IAA17 *cell cycle arrest and release as in Figure 5.9. (**A**) Flow cytometry analysis of samples from the wild-type culture. The majority of cells had completed genome replication by 90 min after the release. (**B**) Flow cytometry analysis of samples from the* pGAL-Os.TIR1, CDC9-IAA17 *strain. The majority of cell in this culture had also reached a DNA content of 2C by 90 min after release, although a proportion of the population still had a DNA content < 2C.*

wild-type cells had completed genome replication and had a DNA content of 2C (Figure 5.13 A). Again, the majority of the *pGAL-Os.TIR1, CDC9-IAA17* population had also reached a DNA content of 2C, although a proportion of the population were yet to complete genome replication (Figure 5.13 B).

Cells harvested 90 min after the G1 release were spheroplasted and embedded in agarose. DNA was purified from the plugs and nicks in the genome were labelled using *Taq* DNA polymerase in a nick-translation reaction in the presence of aminoallyl-dUTP-Atto647, a fluorophore compatible with BNG optical mapping. Translated nicks were then repaired using *Taq* DNA ligase. Next, the labelled and repaired DNA underwent DLE labelling using the BNG kit to label CTTAAG sites in the genome, with an anticipated average density of 13.5 labels/100 kb. Finally, the total DNA was stained according to the BNG DLS kit. This process should have produced molecules suitable for optical mapping with two different labels: aminoallyl-dUTP-Atto647 marking nick sites in the genome following S phase, and DLE providing genomic location labelling, as well as the total DNA stain. To confirm that each of these fluorescence signals could be detected in the treated DNA, samples were analysed by gel electrophoresis and visualised using a variety of filters to capture each labelling and staining step (Figure 5.14). First, an Alexa647 filter was used to detect Aminoallyl-dUTP-Atto647 incorporation at nicks in the DNA (Figure 5.14 A). This showed a band in the *pGAL-Os.TIR1, CDC9-IAA17* samples that was absent in the wild-type samples, consistent with OF labelling in the *pGAL-Os.TIR1, CDC9-IAA17* DNA. An Alexa555 filter was used to visualise the DLE labelling (Figure 5.14 B). The wavelength of the DLE fluorophore was not known, but this filter was chosen based on the lasers used in Saphyr instruments. This showed multiple signal patches for each sample, presumably due to unincorporated fluorophores, since not all patches had corresponding signal in the DNA stain images (Figure 5.14 C + D). The Saphyr DNA stain was presumed to be YOYO-1. Thus a SybrGreen filter was used to visualise this DNA stain, since this filter had the closest overlapping excitation and emission spectra to YOYO-1 of those available. This showed a single band of signal in each of the lanes in which the DNA had been stained (Figure 5.14 C), suggesting successful staining of the molecules. Finally, the gel was treated with

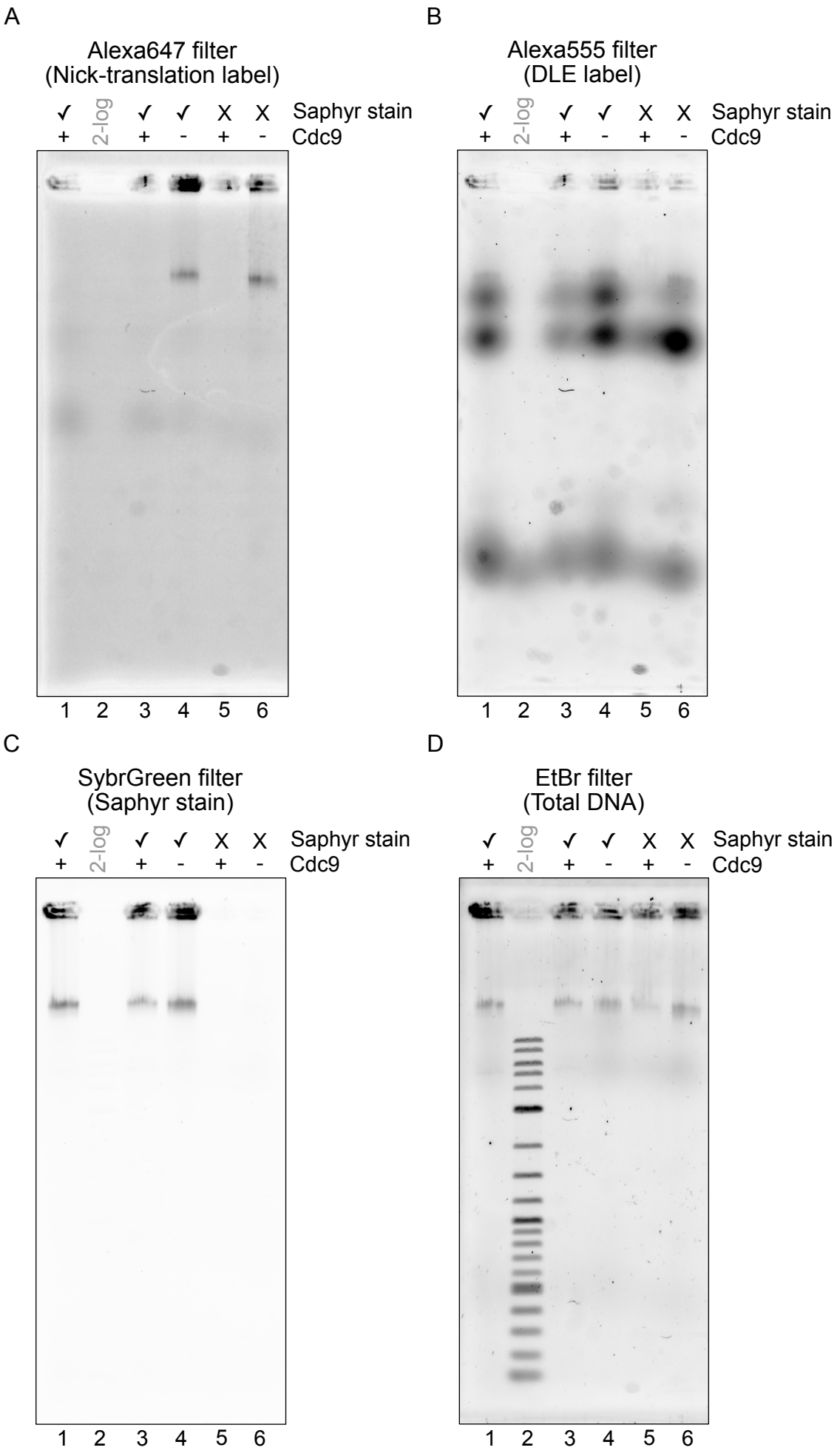

A
Alexa647 filter
(Nick-translation label)
✓ 2-log ✓ ✓ X X Saphyr stain
+ + + - + - Cdc9
1 2 3 4 5 6
B
Alexa555 filter
(DLE label)
✓ 2-log ✓ ✓ X X Saphyr stain
+ + + - + - Cdc9
1 2 3 4 5 6
C
SybrGreen filter
(Saphyr stain)
✓ 2-log ✓ ✓ X X Saphyr stain
+ + + - + - Cdc9
1 2 3 4 5 6
D
EtBr filter
(Total DNA)
✓ 2-log ✓ ✓ X X Saphyr stain
+ + + - + - Cdc9
1 2 3 4 5 6

***Figure 5.14**: Visualisation of* in vitro *DNA labelling and staining by native gel electrophoresis. DNA was isolated from cells harvested 90 min after G1 release in the arrest/release experiment shown in Figure 5.13. DNA nicks were labelled with aminoallyl-dUTP-Atto647 and repaired, prior to DLE labelling and DNA staining. DNA was then size separated by native gel electrophoresis and filters of different wavelengths were used to visualise the DNA labels and stain. Wild-type DNA ("Cdc9 +", lanes 1, 3, and 5) was analysed in addition to* pGAL-Os.TIR1, CDC9-IAA17 *DNA ("Cdc9 -", lanes 4 and 6). Unstained samples were also analysed (lanes 5 and 6). (**A**) Visualisation using an Alexa647 filter to detect aminoallyl-dUTP-Atto647 incorporated during nick translation showed signal exclusively in* pGAL-Os.TIR1, CDC9-IAA17 *samples (lanes 4 and 6). (**B**) Visualisation with an Alexa555 filter to detect DLE labels showed multiple signal bands in all lanes (except the marker in lane 2). (**C**) Visualisation using a SybrGreen filter to detect stained molecules showed signal in samples that had been stained with the BNG DNA stain (lanes 1, 3, and 4). (**D**) Following the visualisations in (A)-(C), the gel was treated with ethidium bromide (EtBr) and then visualised using an EtBr-specific filter. This enabled visualisation of the marker (2-Log) and total DNA in all samples. Each sample had a band of higher molecular weight than the uppermost band in the marker (10 kb).*

ethidium bromide to stain total DNA (Figure 5.14 D). This showed the size marker which indicated that the DNA was of high molecular weight as it had migrated a shorter distance than the uppermost band of the marker (10 kb). Therefore, it was concluded that the DNA labelling and staining reactions were incorporating the two required labels and the stain, while maintaining high molecular weight DNA.

In summary, conditions were found that enabled labelling of nicks in genomic DNA with a fluorescently-labelled nucleotide and subsequent nick repair. In addition, the size of DNA molecules following extraction in plugs and subsequent purification was assessed by PFGE. Finally, it was found that genomic DNA treated to incorporate two labels and the DNA stain required for BNG displayed signal at the anticipated wavelengths for each fluorophore. Thus,

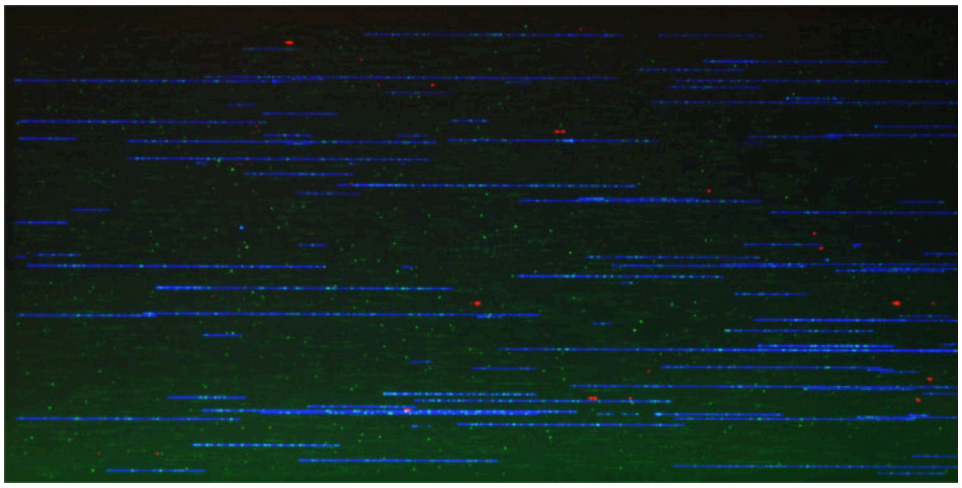
a new protocol for optical mapping of unligated OFs had been checked by gel electrophoresis and was considered to potentially be suitable for Saphyr analysis.

5.5 Optical mapping of *S. cerevisiae* Okazaki fragments using Saphyr

To determine whether the experimental conditions were suitable for optical mapping of OFs using Saphyr, DNA from cells harvested 90 min after the G1 release in Figure 5.13 was labelled as for Figure 5.14. The wild-type and *pGAL-Os.TIR1, CDC9-IAA17* samples were each run in a Saphyr chip flow cell in parallel. Example images from the run indicated low levels of label channel 2 signal (LC2, shown in red), representing the nick label, in the wild-type sample, with higher levels in the *pGAL-Os.TIR1, CDC9-IAA17* sample (Figure 5.15). This matched the expectation of higher levels of unsealed nicks available for labelling in the *pGAL-Os.TIR1, CDC9-IAA17* DNA due to incomplete OF ligation.

The molecule quality reports (MQRs) for each flow cell were used to assess the quality of the data obtained during this run (Table 5.2). This showed that the overall yield was greater for the wild-type than the *pGAL-Os.TIR1, CDC9-IAA17* sample, and that the molecule N50 was also higher, suggesting greater molecule integrity. The percentage of molecules mapping to an *in silico* reference digest (cmap) was surprisingly low, even for the wild-type sample, as this is suggested to fall between 70-90% by BNG. This could be due to single nucleotide polymorphisms (SNPs) between the backgrounds of the strain used (W303) compared to the reference sequence (S288C), but it seems unlikely that this alone would account for the low mapping rate. The negative label variance (defined as the “Percentage of reference labels absent in molecules” in the MQR) was higher than the company’s recommended range of 6-15%. This high proportion of false negatives for the DLE-1 labelling suggested that the reactions did not go to completion and/or that components from the OF labelling reaction interfered with the DLE-1 labelling in some way. Therefore, a potential optimisation of the protocol might be to include a purification step between the nick translation and the DLE-1 labelling steps. This was not included in this trial in order to try and minimise the number of DNA manipulations and therefore preserve maximal molecule integrity. Furthermore, the

A WT



B *pGAL-Os.TIR1, CDC9-IAA17*

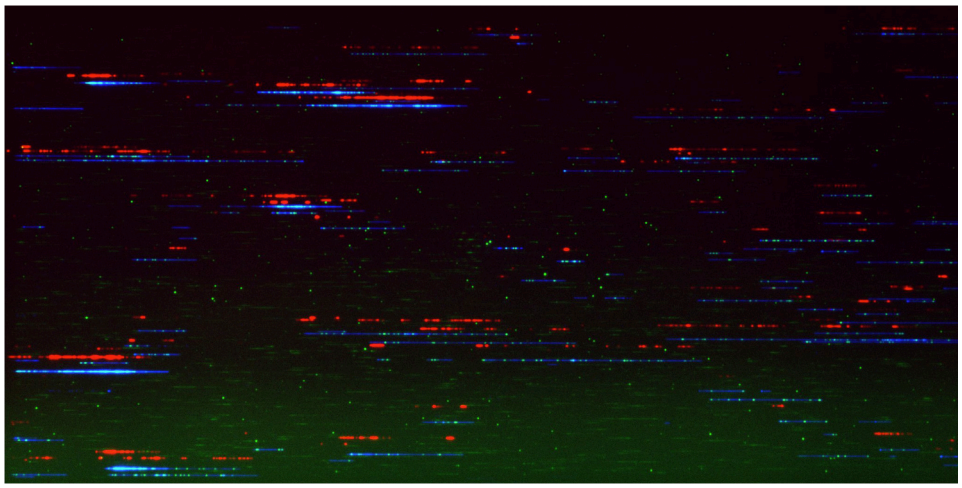


Figure 5.15*: Example images from the Saphyr run. DNA backbone staining is shown by the blue signal (label channel 0 (LC0)), DLE-1 location labelling is shown in green (LC1), and nick-translation labelling is shown in red (LC2). (**A**) Molecules from the wild-type strain. (**B**) Molecules from the* pGAL-Os.TIR1, CDC9-IAA17 strain. *Higher levels of LC2 signal were observed in the* pGAL-Os.TIR1, CDC9-IAA17 *strain compared to the wild-type control. The LC2 signal was offset compared to the LC0 and LC1 signals, which is an established phenomenon caused by chromatic shift* (De Carli et al. 2018).

labelling trial in Figure 5.14 had suggested that all anticipated label wavelengths were present, from which it was inferred that the DLE-1 labelling was not affected by the OF labelling and repair. Interestingly, the average LC2 label density (corresponding to the density of OF labels) was very low in the wild-type sample (0.04/100 kb), therefore giving a rate for the background

level of nick labelling. In contrast, the rate was >250x times higher in the *pGAL-Os.TIR1, CDC9-IAA17* DNA, consistent with much higher levels of label incorporation during nick translation of this sample. This suggested that there was a greater number of nicks to be labelled, given that the two reactions were performed from the same master mix of reagents and so the variation was unlikely to have been caused by technical differences. Therefore, this metric supported the interpretation that unligated OFs were being labelled in this setup.

Table 5.2*: Summary of Saphyr molecule quality metrics for each sample. The metrics correspond to molecules ≥ 150 kb in length. The N50 value refers to the centre of mass of the molecule size distribution. The average DLE-1 density describes the average number of location labels per 100 kb for total molecules ≥ 150 kb, with the anticipation of 13.5 labels per 100 kb based on* in silico *digestion of the reference sequence (sacCer3). The map rate is the percentage of molecules that align to the sacCer3 reference cmap. The positive label variance is the percentage of labels present in molecules but absent in the reference cmap. Conversely, the negative label variance is the percentage of reference labels that were not detected in the molecules. These metric definitions were taken from BNG's Saphyr Molecule Quality Report Guidelines (Document Number 30223, Revision B).*

Sample	Total DNA (Gb)	N50 ≥ (kb)	Average DLE-1 label density	Map rate (%)	Effective coverage	Positive label variance (%)	Negative label variance (%)	Average LC2 label density
Wild-type	481	274	11.61/100 kb	53.2	20,950x	4.7	22.6	0.04/100 kb
pGAL-Os.TIR1, CDC9-IAA17	172	193	12.80/100 kb	14.0	1,974x	3.6	25.2	10.37/100 kb

Given that this was a novel technique, an established pipeline for data analysis was not available. Furthermore, the new Saphyr system generates data that is not backward compatible with software written to analyse Irys data (Arielly and Ebenstein 2018). The raw

molecule information from a Saphyr run is outputted in a bnx file, which contains information about the label locations and quality scores. To analyse the positions of the LC2 signals across the genome, it was first necessary to align the bnx file to the reference cmap using BionanoSolve. This produced cmap files which contained information about the detected labels within each molecule, relative to that molecule (i.e. not in relation to the reference sequence), and xmap files which provided information about the alignment between each molecule and the reference. The next step was to combine the data contained within the different files to obtain positional information about each label in relation to the reference genome. Initially, this was achieved using a custom R script kindly written by Phoebe Oldach. This script linked aligned molecules in cmap and xmap files to enable the calculation of each label's genomic coordinate, based on its distance from the molecule start and the molecule's position within the genome. The resulting output in bedgraph format was visualised in a genome browser (Figure 5.16). This showed LC2 label incorporation along the full length of molecules to varying extents. The signal intensity of each label also varied considerably.

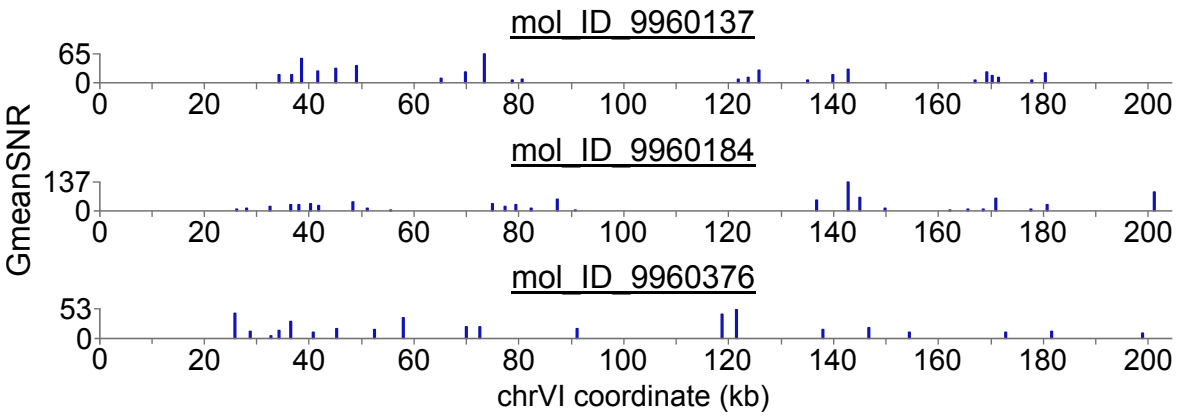


***Figure 5.16**: Three example molecules aligned to chromosome VI. The LC2 channel signal is shown as blue bars, with the height representing the GmeanSNR of each label (note the different y-axis scales for each molecule). The total length of each molecule was as follows: 9960137 = 180,518 bp, 9960184 = 192,868 bp, and 9960376 = 204,871 bp.*

From the LC2 positional information, it was possible to calculate the distance between each signal and its neighbour within a molecule. This was considered to be potentially useful when

determining how to group adjacent LC2 signals as a tract of signal, rather than an individual point. The distribution of these distances showed that the majority of LC2 signals were within 10 kb of one another (Figure 5.17), with a median distance of 3,831 bp between neighbours. Although the OF length distribution has been shown to reflect the 165 bp nucleosomal repeat length (Smith and Whitehouse 2012), the inherent resolution limit of optical mapping would have prevented the resolution of individual OF labels within 1 kb in these data. Furthermore, any residual DNA ligase activity during S phase would have increased the distances between adjacent nick sites. However, the skew of the distances suggested a non-random distribution of labelled sites throughout the genome and was consistent with label clusters interspersed with longer stretches without label. This matched the expected pattern of regions of lagging strand alternating with regions of leading strand synthesis along each molecule.

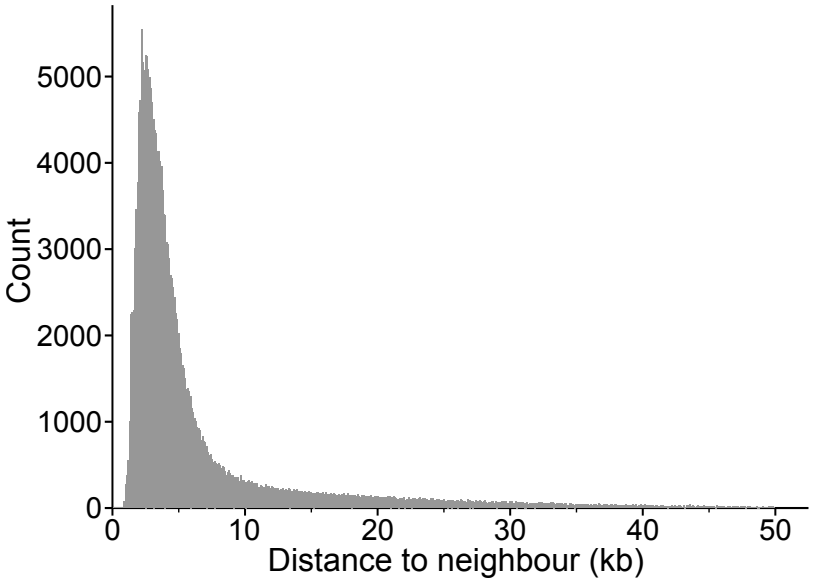


Figure 5.17*: Histogram showing the distribution of distances between neighbouring LC2 labels in the* pGAL-Os.TIR1, CDC9-IAA17 *sample plotted in 100 bp bins. The median distance was 3,831 bp. Very few distances below 1 kb were measured and there was a rapid decline in the counts per bin above approximately 3 kb. For visualisation purposes, the x-axis of the plot was limited to 50 kb.*

To assess whether there was a bias in the signal with respect to replication time, the LC2 signals were compared to sort-seq data (Müller et al. 2014). It was postulated that OFs

synthesised in early S phase would have longer exposure to any residual DNA ligase activity than those synthesised later, and thus may have a lower labelling density. To check whether the distribution in LC2 signal across S phase was constant or biased, the genome was divided into 10 bins depending on the measured DNA copy number in S phase by sort-seq (Müller et al. 2014). The overlap between the positions of all detected LC2 signals in the aligned *pGAL-Os.TIR1, CDC9-IAA17* sample and these different sort-seq bins was compared (Figure 5.18).

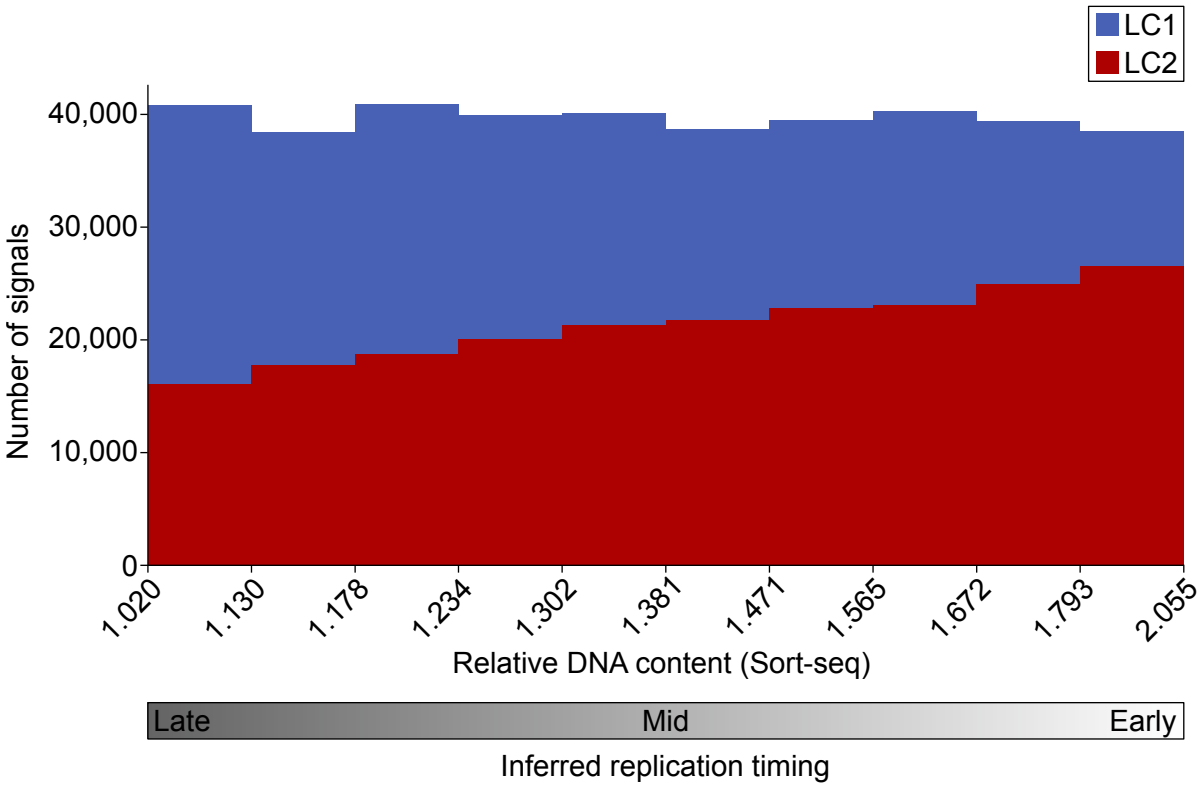


Figure 5.18*: Distribution of LC1 and LC2 signals with respect to genome replication timing. Sort-seq data* (Müller et al. 2014)*, providing a relative DNA content between 1 and 2 in 1 kb windows in the genome, were split into deciles based on the relative DNA content. The intersect between each of the deciles and the genomic location of each LC1 or LC2 signal was determined using bedtools intersect. The number of LC1 (blue) and LC2 (red) signals intersecting with each DNA content decile was plotted. High relative DNA content in sort-seq data indicates early replication. The number of LC2 signals intersecting with each DNA content decile increased as the relative DNA content increased.*

This showed an apparent decrease in LC2 signals as the replication timing became later, contrary to the expectation. A possible explanation for this observation was the presence of cells still undergoing DNA replication in the population, as evidenced by the analysis of total DNA content by flow cytometry (Figure 5.13). However, this was not investigated further and additional analyses would be required to address the hypothesis given above.

It was subsequently found that BNG had written a custom R script to calculate the chromosomal coordinate of each LC2 signal. This script was kindly shared by Dr. McGowan (University of Oxford) and used to obtain the genomic coordinates of LC2 signals in the *pGAL-Os.TIR1, CDC9-IAA17* sample. To compare the outputs of the in-house and BNG scripts, the same three example molecules were plotted as in Figure 5.16. In addition, an overlay of the results of the two scripts was plotted to facilitate comparisons (Figure 5.19). This showed near

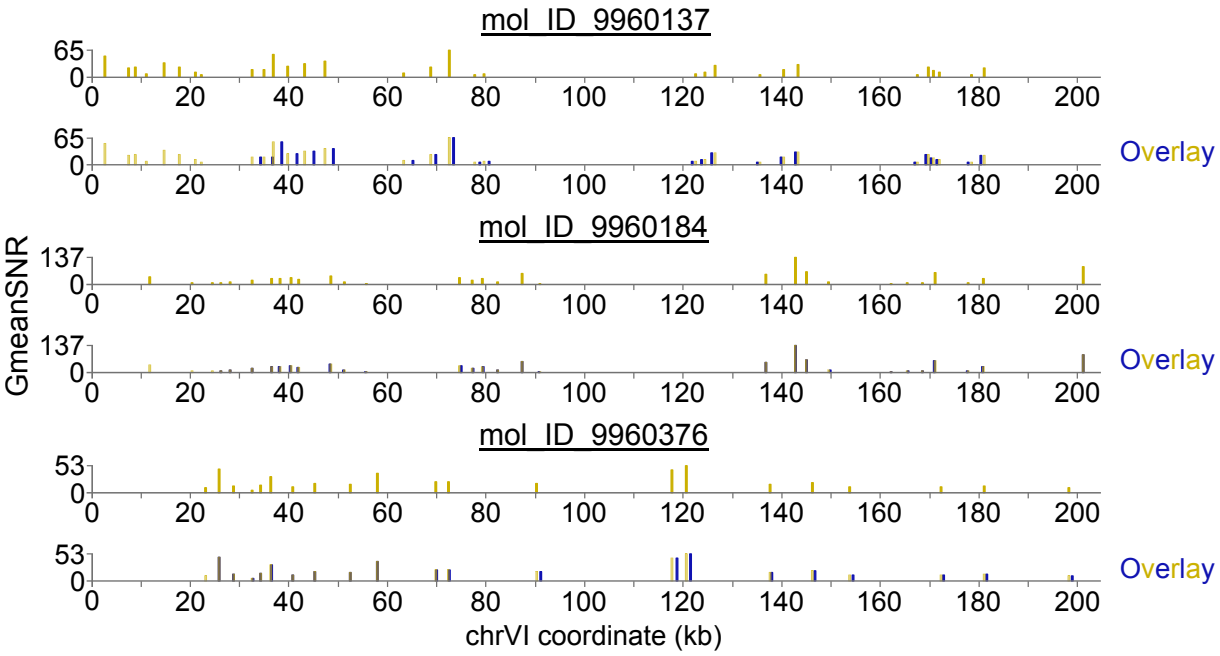


***Figure 5.19**: The same three example molecules as shown in Figure 5.16 were plotted following processing by the BNG script. For each molecule, the signal positions of the LC2 channels are shown in the top track by the yellow bars. The track below each of these shows an overlay of the two script outputs, with the results from the in-house script shown in blue and the results from the BNG script shown in yellow with reduced opacity. Examples of both concurrence and divergence between the outputs were observed.*

complete agreement for some molecules (e.g. 9960184), while other molecules showed minor offsets in label positions (e.g. 9960137). Furthermore, the BNG script detected the presence of LC2 signals in some regions that were not detected with the in-house script (e.g. signals below 30 kb in molecule 9960137). The reasons for these discrepancies were not investigated. However, they could be due to the inclusion of bnx data as an input in the BNG script (only the cmap and xmap data were used for the in-house script).

Having processed the raw data to enable alignment of single molecules to the reference and subsequent mapping of LC2 label positions, it was then necessary to assess whether replication information could be extracted from the data. To qualitatively determine whether one end of each cluster of LC2 signals aligned with a known origin of replication, single molecules were plotted as in Figures 5.16 and 5.19, following processing by the in-house script and compared to annotated replication origins (Figure 5.20 A). This showed that the end of each cluster of LC2 signals was frequently in close proximity to a known origin. For example, on molecule 5048229 the left-ends of each cluster of LC2 signal are close to *ARS603.5*, *ARS606* and *ARS607*, consistent with these origins being responsible for the replication of this molecule. By comparison, on molecule 9299326 the right-ends of each cluster of LC2 signal are close to replication origins (*ARS603*, *ARS603.5* and *ARS606*). A possible approach for automating the analysis of replication dynamics could be to assign tracts of lagging strand synthesis based on the distances between neighbouring LC2 signals (conceptually illustrated for molecule 9299326 in Figure 5.20 B). This would require the determination of a threshold distance above which neighbouring signals would be assigned as belonging to separate tracts. The lagging strand tracts could then be used to systematically infer the location of replication initiation events (by comparison to known origins), the direction of replication forks and therefore the relative replication time across each molecule (Figure 5.20 C) (Retkute, Nieduszynski, and De Moura 2012).

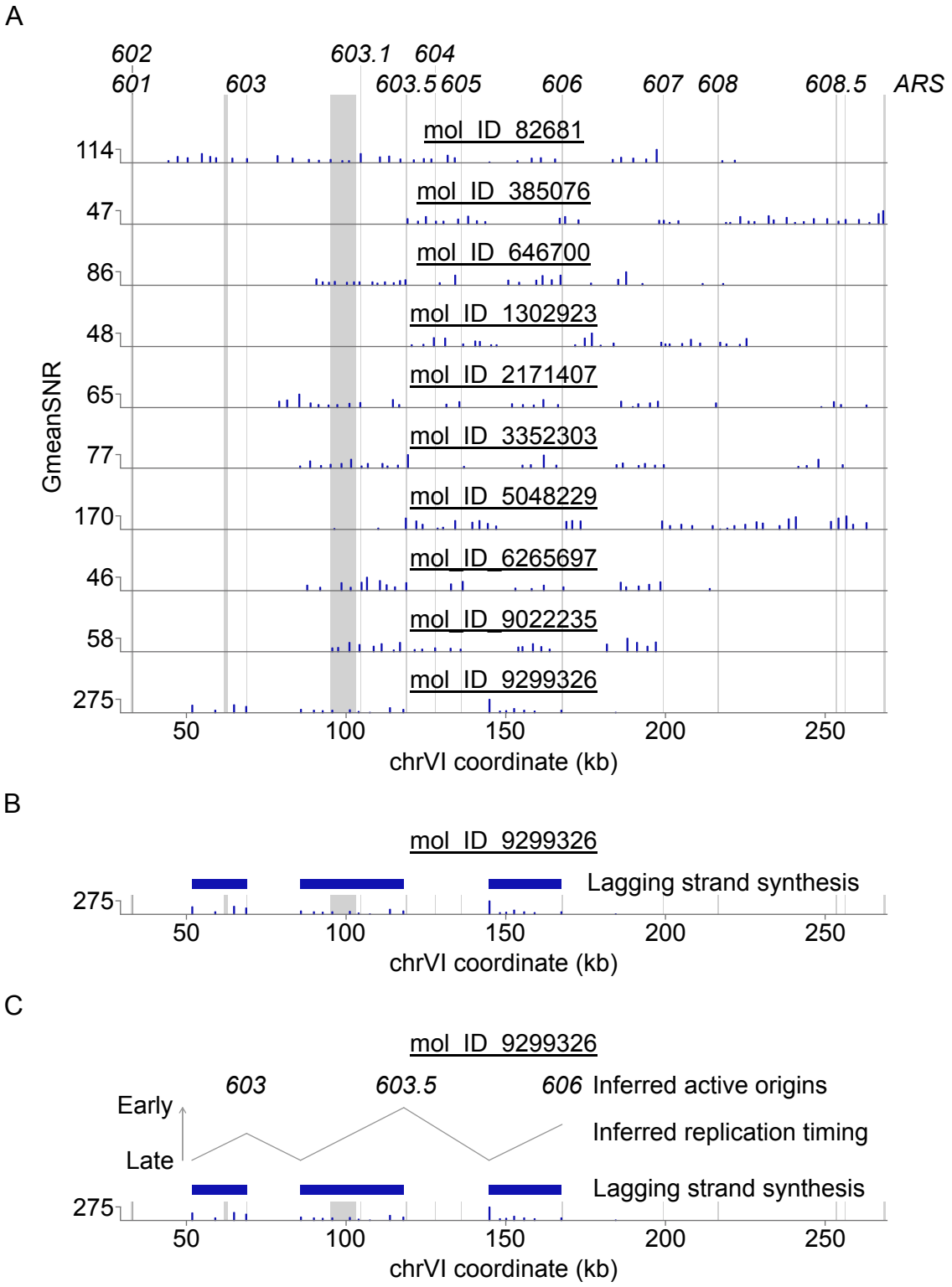


***Figure 5.20**: Comparison of single molecule LC2 signals with known replication origins. (**A**) Ten example molecules were mapped to chromosome VI using the in-house script. The grey bars represent confirmed and likely origins from OriDB* (Siow et al. 2012)*. The height of each LC2 bar corresponds to its GmeanSNR. (**B**) Theoretical assignment of tracts of lagging strand synthesis based on the sites of LC2 signal in an example molecule. (**C**) The theoretical tracts in (B) could be used to ascertain which origins were active in the single molecule. The length*

of each tract could be used to infer the relative timing of each origin, if a constant fork speed is assumed.

An existing optical mapping study described the assignment of tracts of label incorporation in human cells pulsed with fluorescently-labelled dUTP (Klein et al. 2017). To determine whether this approach was appropriate for the assignment of OF tracts in *S. cerevisiae*, the alignment data were analysed using the script written for the above study (Figure 5.21). Running the script with the same thresholds as described by the authors for grouping LC2 signals as tracts (script edit 1) grouped apparently distinct tracts into one large cluster, suggesting that these parameters were suboptimal for the analysis of these data. The thresholds were varied to assess the impact on tract calling (script edits 2-5). Edits 4 and 5 qualitatively appeared to best

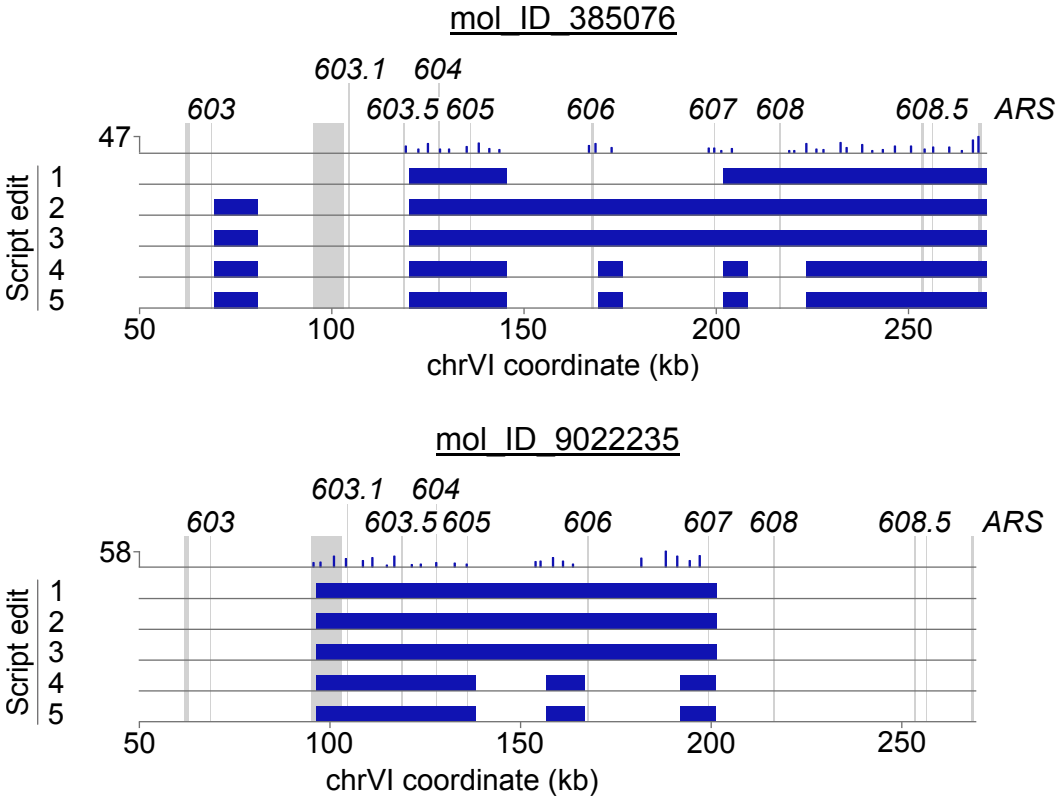
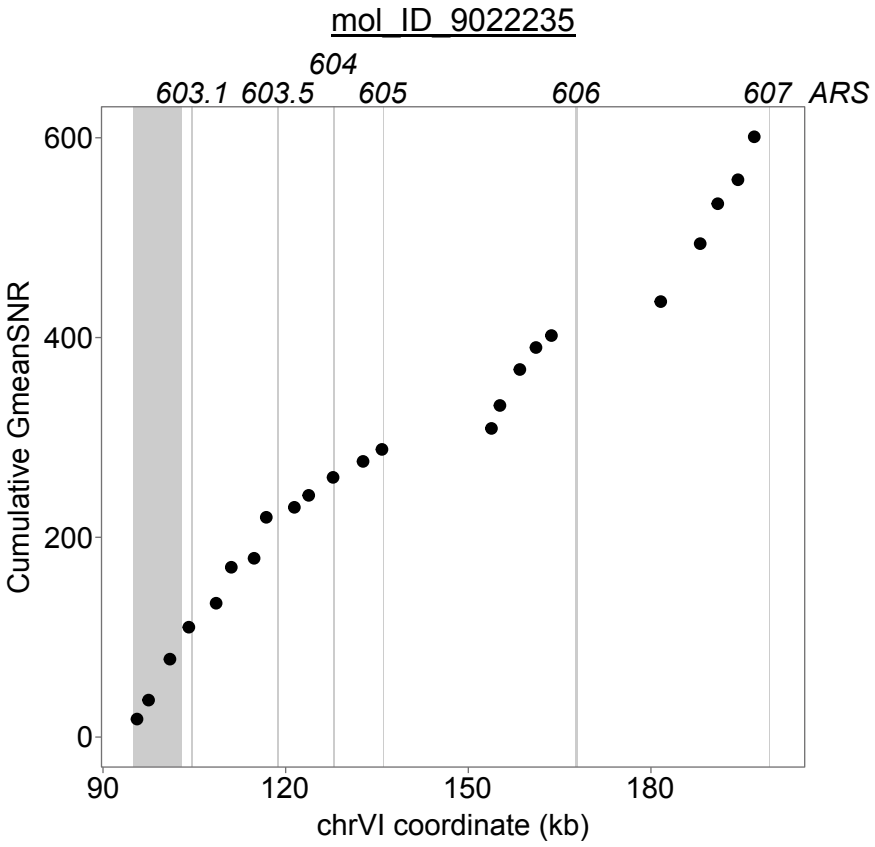


***Figure 5.21**: Visualisation of tracts called using a publicly-available script (https://github.com/tborrman/DNA-rep/tree/master/BioNano). The parameters of the script were edited (see section 2.8.3) in an attempt to identify the optimal conditions for the assignment of lagging strand tracts. The different outputs were plotted along with the underlying LC2 signals (top track in each molecule).*

represent the underlying LC2 signals. These edits reduced the number of proximal signals required to define a tract, as well as reduced the distance between neighbouring labels. However, further optimisation or subsequent analysis of the called tracts was not performed.

An alternative approach to mapping single molecule replication dynamics could involve the cumulative summing of the GmeanSNR of LC2 signals across a molecule and detection of where gradient changes occur. The cumulative GmeanSNR was found to show gradient transitions between a positive slope and a flat line for two example molecules (Figure 5.22). The regions of no increase in the cumulative GmeanSNR can be inferred to have been synthesised as the leading strand, while the regions with a changing cumulative GmeanSNR would represent lagging strand synthesis (theoretical assignment shown for molecule 385076). This would then allow the inference of initiation events in each molecule (e.g. *ARS603.5*, *ARS606, ARS607* and *ARS608* in molecule 385076).

A


mol_ID_9022235
604
603.1 603.5 605 606 607 ARS
600
400
200
0
Cumulative GmeanSNR
90 120 150 180
chrVI coordinate (kb)

B

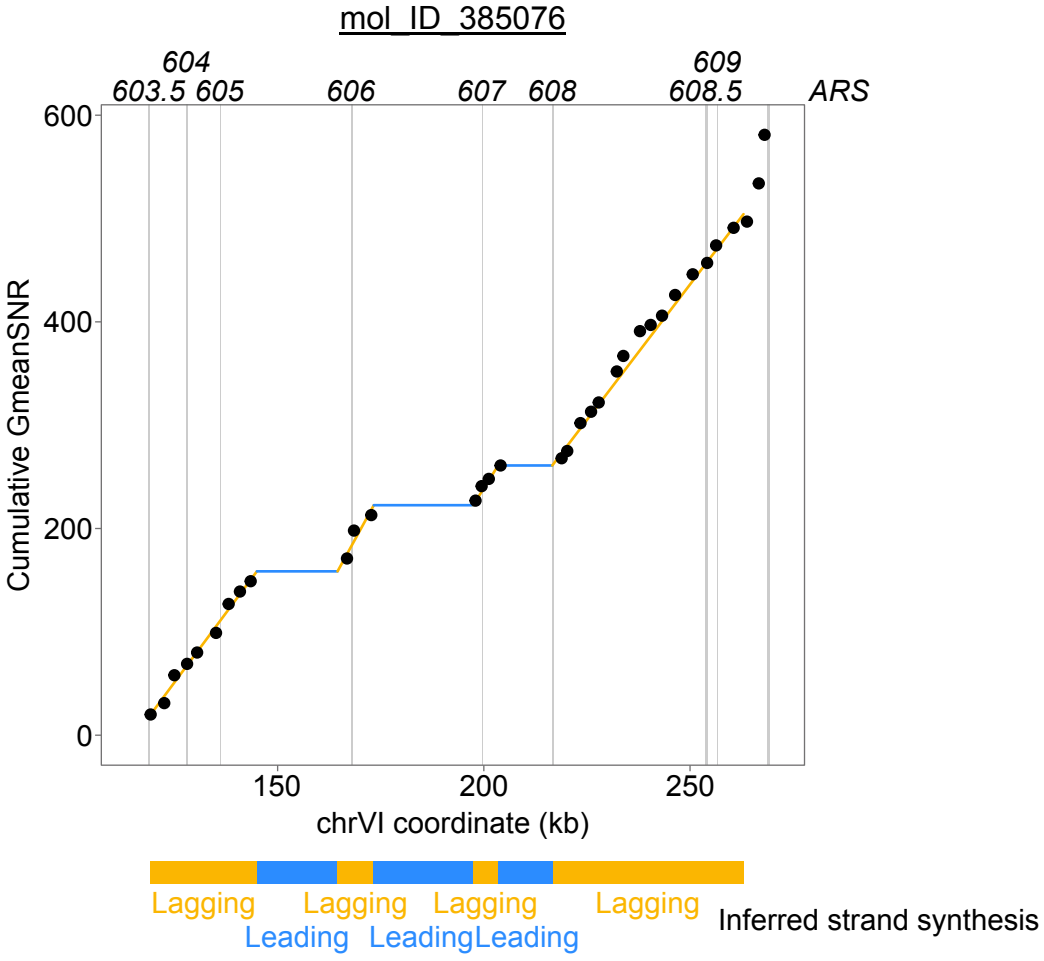

mol_ID_385076
604
603.5 605 606 607 608 609 608.5 ARS
600
400
200
0
Cumulative GmeanSNR
150 200 250
chrVI coordinate (kb)
Lagging Lagging Lagging Lagging
Leading Leading Leading
Inferred strand synthesis

***Figure 5.22**: Cumulative GmeanSNR along two example molecules. The cumulative GmeanSNR of LC2 signals was calculated along the each molecule and plotted relative to the chromosomal coordinate. Known origins are shown as grey bars. The plot in (**B**) was manually annotated with potential lines of best fit for the different segments of the data. If an approach were developed to calculate this, it would allow the regions synthesised as leading and lagging strand to be annotated as shown beneath the plot.*

5.6 Summary

In summary, preliminary analysis indicated that nicks were specifically labelled in DNA from the *pGAL-Os.TIR1, CDC9-IAA17* strain, consistent with the labelling of unligated OFs. It was necessary to manipulate the raw data to determine the chromosomal coordinate of each label, and an in-house script was developed for this purpose. This enabled the LC2 signals of each molecule to be visualised in a genome browser in relation to annotated replication origins. Initial investigation into the assignment of replication dynamics in single molecules suggested that the molecules were showing the anticipated pattern of alternating leading and lagging strand synthesis, with transitions occurring proximal to known origins. However, further analysis would be required to validate these approaches and to reveal the genome-wide replication dynamics on single molecules.

Chapter 6: Discussion

The work summarised in this thesis aimed to expand the repertoire of available single molecule methods for the study of genome replication dynamics. Two main projects were pursued to this end: first, the adaptation of a recently-developed method to *S. pombe*, and second, the development of a novel optical mapping technique to detect differences in leading and lagging strand synthesis by one of two approaches.

Genome replication dynamics are an important factor in preserving genome integrity (Yoshida, Poveda, and Pasero 2013) and in the regulation of certain cellular processes, such as gene expression level (Müller and Nieduszynski 2017). While this has been extensively studied in a wide range of cell types and species using a plethora of population-based techniques, a recurring challenge has been the elucidation of the variability in the process between cells of the same population and during developmental changes in metazoan cells (Rivera-Mulia and Gilbert 2016). There has been, therefore, a long-established requirement for methods that enable the investigation of heterogeneity in replication dynamics at the single cell and single molecule level. For several decades, DNA fibre analysis and DNA combing have been used to visualise ultra-long individual molecules by fluorescence microscopy (Czajkowsky et al. 2008; Weier 2001; Bianco et al. 2012). Such techniques have been applied to the study of replication dynamics through the incorporation of nucleotide analogues by the cellular replication machinery. In addition to the single molecule-nature of such approaches, a further advantage is the interrogation of replication patterns in ultra-long molecules, even entire *S. pombe* chromosomes (Kaykov et al. 2016). This provides insight that can only be inferred, rather than directly investigated, using short-read, population-based approaches. While it is difficult to overstate the value of such methods, a number of limitations reduce their suitability for certain analyses. First, DNA combing requires specialist equipment and is a technically challenging method. Second, molecules are predominantly anonymous, unless sequence-specific probes are used (which cannot be applied genome-wide). Third, it is not a high-throughput, automated process, so laborious data acquisition and analysis are involved,

reducing the potential coverage of the method. These last two shortcomings are addressed by the two single molecule approaches explored in this work.

Bionano Genomics (BNG) platforms combine genome-wide optical mapping with automated data acquisition to produce high-coverage genomic datasets, comprising ultra-long single molecules. Oxford Nanopore Technologies (ONT), on the other hand, provide single-molecule sequencing methodologies, also yielding high-genome coverage by high molecular weight molecules.

The first project described in this work involved the application of the recently-described D-NAscent technique to *S. pombe*. This methodology was developed in *S. cerevisiae* (Müller et al. 2019), which differs in some crucial experimental aspects to *S. pombe*. For instance, *S. cerevisiae* strains of the *MATa* genotype are routinely arrested in G1 phase by incubation with the mating pheromone, α-factor. While *S. pombe* strains are similarly susceptible to treatment with pheromone from cells of the opposite mating type (Nielsen 2016), this is not generally used for cell cycle synchronisation in this species. More commonly, cell-cycle arrest and release is controlled in temperature-sensitive (ts) mutants which can be arrested in a particular cell cycle phase by incubation at their restrictive temperature (Hagan, Grallert, and Simanis 2016). For this work, a *cdc10-m17* ts mutant was used, which can be arrested in G1 phase (Nurse, Thuriaux, and Nasmyth 1976), then released at its permissive temperature for synchronous progression through S phase. An important caveat of this arrest/release approach is the perturbation from a “normal” S phase following incubation at the restrictive temperature. Indeed, cell cycle synchronisation in G2 phase using a *cdc25-22* ts mutant has previously been shown to alter replication dynamics when compared with use of elutriation for enriching for cells in G2 phase (Xu et al. 2012). Although a different ts mutant was used in this work, it is likely to deviate from replication dynamics in wild-type cells. At the very least, the cellular context of replication is altered from the usual conditions of a binucleate cell to cells that have completed cytokinesis during the G1 arrest, before entry into S phase. The impact of this on replication dynamics was not investigated, but it could reasonably be assumed that

there would be some effect which would have to be taken into account during any interpretations of the data. Alternative synchronisation approaches were considered, in particular enrichment of early G2 phase cells by elutriation or lactose gradients. However, in both cases, the anticipated yield of synchronous cells was deemed to be too low for subsequent D-NAscent analysis. Furthermore, elutriation requires specialist equipment and technical expertise that were not readily available, while lactose gradients were not thought to yield as high levels of synchrony as the alternative approaches (Prof. T. Carr, personal communication).

Successful cell cycle synchronisation in the *cdc10-m17, adh:hENT1, adh:hsv-TK* strain along with addition of BrdU 20 min prior to release from the G1 arrest was shown to result in genome-wide incorporation of BrdU in nascent DNA. However, consistent with the perturbation of the usual replication timing programme in such an experimental setup, the BrdU signal was fairly uniform across the genome. Additional arrest and release in hydroxyurea (HU) and a thymidine chase were required to localise BrdU incorporation to only the earliest-replicating regions. However, the detected levels were insufficient for reliable predictions about single molecule replication dynamics in this experimental setup, indicating that further optimisation was required. The desired outcome of such optimisations would be high levels of BrdU incorporation specifically in early S phase, at early-firing origins, which would then gradually decrease to background levels as BrdU was outcompeted by thymidine. Such gradients in BrdU signal, marking transitions from high to low levels of incorporation, have already been shown to allow the reliable detection of fork direction and therefore replication initiation events in *S. cerevisiae* (Müller et al. 2019). Thus, the goal would be to achieve such patterns of BrdU incorporation in *S. pombe*. Work performed by Dr. C. Müller in continuation of the original D-NAscent study has indicated that the number of copies of the thymidine kinase incorporated into the genome of experimental *S. cerevisiae* strains has an impact on their suitability for D-NAscent analysis, with just one copy being suboptimal. It may be that there is a similar dependence in *S. pombe* strains. A genome assembly of the nanopore data produced during this work, kindly performed by Dr. Müller, confirmed that the *cdc10-m17, adh:hENT1, adh:hsv-*

TK strain indeed has one copy of the kinase. Therefore, to further optimise the protocol in *S. pombe*, a potential next step could involve genetic modification to introduce two or more copies of the kinase into the genome of the experimental strain. Furthermore, the identification of the equilibrative nucleoside transporter (ENT) small molecule inhibitor, nitrobenzylthioinosine (NBMPR), following a review of the literature by Dr. Müller (Boleti et al. 1997) raised the interesting possibility of greater control over the import of BrdU from the medium. This could theoretically limit BrdU uptake to any desired cell cycle phase by the experimentalist, and offers an exciting future direction for further optimisation of the technique.

In the second project described in this thesis, a novel method involving BNG analysis was investigated. In the first instance, the goal was to generate optical maps of polymerase usage based on methods exploiting ribonucleotide incorporation during replication (Reijns et al. 2015; Koh et al. 2015; Clausen et al. 2015; Daigaku et al. 2015). Attempts to generate budding or fission yeast strains suitable for this approach proved unsuccessful and the focus was turned to Okazaki fragment (OF) detection. Through depletion of DNA ligase, the enzyme responsible for the ligation of adjacent OFs, the unresolved nicks in the DNA backbone of nascent strands could be labelled by nick translation. A procedure for obtaining such labelled DNA was developed and assessed by Saphyr. Preliminary analysis of the resulting data was consistent with successful OF labelling specifically in the DNA ligase-depleted strain. However, the development of a thorough data analysis pipeline was beyond the scope of this project, so a number of further analyses are required to further validate the method. A crucial aspect of this will be the generalisation and automation of an approach for mapping the replication dynamics in individual molecules. Based on the current data, two potential solutions to this problem are proposed here. The first of these requires tract calling based on the distribution of signals along a molecule. Labels separated by a distance below an appropriate threshold would be clustered into a tract, while those further than the threshold distance would be assigned as a separate cluster (or an individual signal, depending on the proximity of neighbouring signals). The transitions from a cluster of signal to no signal would thus demarcate transitions between lagging and leading strand synthesis. Due to the nature of the data, it would not be known

whether such transitions represented an initiation or termination event. However, given the known initiation sites in the *S. cerevisiae* reference genome, alignment of such tracts to the reference should enable the determination of replication direction in each molecule based on the location of tract ends in relation to the known sites. The second possible approach for extracting replication dynamics information from the molecules would be to cumulatively sum the intensity of the detected OF signals and plot this with respect to the label coordinates. The anticipation would be a sloping gradient where lagging strand synthesis took place, as the label incorporation should increase in such regions, thereby adding to the cumulative score. Conversely, in regions of leading strand synthesis, where only background levels of label are anticipated, the cumulative score should not increase, and therefore the line would flatten. It would be necessary to develop a machine-learning-based approach to assign the leading and lagging strand regions within each molecule, which it would be necessary to train on a subset of manually-analysed molecules. The importance of the signal-to-noise ratio of each label is currently unclear, so the dependence of this approach on these values would need to be assessed in terms of its suitability.

Once the data analysis pipeline has been further developed, the approach could be validated by comparison with other replication datasets from budding yeast, both at the population and single molecule level. D-NAscent (Müller et al. 2019) provides complementary data to the OF optical mapping data, therefore making the available datasets valuable for comparison with the OF data. The two methods use different approaches for detecting active replication – D-NAscent involves a pulse of analogue incorporation during early S phase, while OF optical mapping marks sites of lagging strand synthesis throughout S phase. Furthermore, an alternative, recently-published method for the detection of BrdU incorporation in single strands by nanopore sequencing has been described (Hennion et al. 2020). This involved a brief BrdU pulse to asynchronously growing cells, and so allowed replication dynamics throughout S phase to be investigated. However, a challenge with this is that the progress of each cell through S phase in an asynchronous population cannot be linked to the single molecules being analysed for replication dynamics. Therefore, the timing of each initiation or termination event

in each molecule cannot be determined. Despite this, the FORK-seq data provide another exciting and insightful dataset to which the OF optical mapping data can be compared and therefore validated.

While it has not yet been fully validated, the OF labelling approach raises the possibility of exciting future work. A unique feature of the dataset is the complete record of how each molecule was replicated during S phase, rather than a snapshot of a particular time. This should enable the inference of how far each replication fork travelled, and whether common fork stalling events can be detected by asymmetry between the distances synthesised by forks emanating from the same initiation event. Furthermore, the replication of long repetitive regions that are difficult to study with short-read methods, such as the rDNA array, could be explored. Therefore, further analysis of the current dataset should enable investigation into outstanding questions about the dynamics of genome replication in budding yeast. For example, the ultra-long single molecules can be analysed to assess whether or not proximal origins activate in a coordinated manner, or whether they fire independently. In addition, the origin properties of timing and efficiency could be disentangled by determining whether origins that have been described as late-firing based on population level-data are indeed late firing, or whether their supposed lateness is in fact due to low efficiency of firing in the population. The high coverage on very long molecules provided by the optical mapping data enables the interrogation of such questions about efficiency that are not readily answered by lower-coverage data.

Possible future work could include the combination of the required genotype for OF mapping with a mutation of interest to explore the impact of that mutation on the replication dynamics in single molecules. However, an important caveat to this would be a potential interaction between the *pGAL-Os.TIR1, CDC9-IAA17* genotype and a mutation of interest to produce a synthetic phenotype. Therefore, it would be necessary to assess this possibility for every genetic combination used. Alternatively, or in addition, the method could be translated into other eukaryotic systems, including human cells, where it has recently been demonstrated that sequencing following siRNA-mediated depletion of DNA ligase 1 in a DNA ligase 3 knockout

cell line can reproduce replication profiles obtained using established methods (Sriramachandran et al. 2020). Therefore, depletion of DNA ligase in such cells, followed by careful extraction of high molecular weight DNA would allow the labelling of unligated OFs and therefore the determination of single molecule replication dynamics. Given that replication initiation sites have not been well characterised in other eukaryotes, this could provide important insight into the pattern of site usage and inter-site distances in a species or cell-type of interest. Furthermore, the impact of treatment with exogenous agents, such as chemotherapeutic drugs, on replication could be characterised.

In summary, this work has described the investigation into two approaches for the study of replication dynamics in single molecules. As mentioned above, the outputs of the two approaches are distinct and complementary datasets. While D-NAscent yields high genome coverage by high molecular weight single strands, the molecule lengths cannot yet match those achieved using the extraction protocol involved in the optical mapping approach. On the other hand, the inherent resolution limit associated with optical mapping reduces the resolution of any associated techniques, while the recent advances in the DNAscent_v2 software enable the detection of BrdU at each thymidine position (Boemo 2020). Therefore, the two methodologies have particular advantages and limitations which influence the suitability of the technologies to different biological questions. Furthermore, the approaches to detect replication dynamics by the two different readouts differ in the nature of the data. The approach described here for optically mapping OFs across the genome is expected to report the replication dynamics for the entire S phase, while techniques involving analogue pulses are limited to providing insight into synthesis that occurred during the pulse. Therefore, with the current experimental setup, D-NAscent focusses particularly on early S phase. This provides clear evidence about early firing origins of replication (Müller et al. 2019), which could only be inferred from the OF approach. Therefore, moving forwards, it will be important to compare and corroborate observations and conclusions from the different datasets.

To conclude, for both projects, more thorough analysis is required to yield biological insight from the novel datasets. However, the current analyses suggest that the optical mapping of lagging strand synthesis may be useful for the study of replication dynamics in long single molecules. Following optimisation and validation of the protocol, a wide range of biological questions could be investigated from budding yeast to metazoan cells. For example, characterisation of replication dynamics in human cells over the course of an entire S phase. In addition, the impact of both endogenous and exogenous replication stresses could be examined using both approaches, once optimised. This could be important for the understanding of alterations in replication dynamics during normal development in metazoa, or in the variations in the replication timing programme observed in cancer progression.

References

Arias, Emily E., and Johannes C. Walter. 2007. “Strength in Numbers: Preventing Rereplication via Multiple Mechanisms in Eukaryotic Cells.” *Genes and Development* 21 (5): 497–518.

Arielly, Rani, and Yuval Ebenstein. 2018. “Irys Extract.” *Bioinformatics* 34 (1): 134–36.

Ayyagari, Rao, Xavier V. Gomes, Dmitry A. Gordenin, and Peter M.J. Burgers. 2003. “Okazaki Fragment Maturation in Yeast: I. Distribution of Functions between FEN1 and DNA2.” *Journal of Biological Chemistry* 278 (3): 1618–25.

Bähler, Jürg. 2005. “Cell-Cycle Control of Gene Expression in Budding and Fission Yeast.” *Annual Review of Genetics* 39 (October): 69–94.

Balakrishnan, Lata, and Robert A. Bambara. 2013. “Okazaki Fragment Metabolism.” *Cold Spring Harbor Perspectives in Biology* 5 (2): a010173.

Bardwell, Lee. 2005. “A Walk-through of the Yeast Mating Pheromone Response Pathway.” *Peptides* 26 (2): 339–50.

Basi, Gabriele, Elisabeth Schmid, and Kinsey Maundrell. 1993. “TATA Box Mutations in the Schizosaccharomyces Pombe Nmt1 Promoter Affect Transcription Efficiency but Not the Transcription Start Point or Thiamine Repressibility.” *Gene* 123 (1): 131–36.

Bell, Stephen P., and Bruce Stillman. 1992. “ATP-Dependent Recognition of Eukaryotic Origins of DNA Replication by a Multiprotein Complex.” *Nature* 357 (6374): 128–34.

Bianco, Julien N., Jérôme Poli, Julie Saksouk, Julien Bacal, Maria Joao Silva, Kazumasa Yoshida, Yea Lih Lin, Hélène Tourrière, Armelle Lengronne, and Philippe Pasero. 2012. “Analysis of DNA Replication Profiles in Budding Yeast and Mammalian Cells Using DNA Combing.” *Methods* 57 (2): 149–57.

Blow, J. Julian, and Anindya Dutta. 2005. “Preventing Re-Replication of Chromosomal DNA.” *Nature Reviews Molecular Cell Biology* 6 (6): 476–86.

Bøe, Cathrine A., Ignacio Garcia, Chen Chun Pai, Jeffrey R. Sharom, Henriette C. Skjølberg, Erik Boye, Stephen Kearsey, Stuart A. MacNeill, Michael D. Tyers, and Beáta Grallert. 2008. “Rapid Regulation of Protein Activity in Fission Yeast.” *BMC Cell Biology* 9 (23).

Boemo, Michael A. 2020. “DNAscent v2 : Detecting Replication Forks in Nanopore Sequencing Data with Deep Learning.” *BioRxiv*, 2020.11.04.368225.

Boettcher, Barbara, and Yves Barral. 2013. “The Cell Biology of Open and Closed Mitosis.” *Nucleus* 4 (3): 160–65.

Boleti, Haralabia, Imogen R. Coe, Stephen A. Baldwin, James D. Young, and Carol E. Cass. 1997. “Molecular Identification of the Equilibrative NBMPR-Sensitive (Es) Nucleoside Transporter and Demonstration of an Equilibrative NBMPR- Insensitive (Ei) Transport Activity in Human Erythroleukemia (K562) Cells.” *Neuropharmacology* 36 (9): 1167–79.

Budd, Martin E., and Judith L. Campbell. 1995. “A Yeast Gene Required for DNA Replication Encodes a Protein with Homology to DNA Helicases.” *Proceedings of the National Academy of Sciences of the United States of America* 92 (17): 7642–46.

Burgers, Peter M.J. 1998. “Eukaryotic DNA Polymerases in DNA Replication and DNA Repair.” *Chromosoma* 107 (4): 218–27.

Burgers, Peter M.J., Dmitry Gordenin, and Thomas A. Kunkel. 2016. “Who Is Leading the Replication Fork, Pol ε or Pol δ?” *Molecular Cell* 61 (4): 492–93.

Caddle, M S, and M P Calos. 1994. “Specific Initiation at an Origin of Replication from Schizosaccharomyces Pombe.” *Molecular and Cellular Biology* 14 (3): 1796–1805.

Carlson, Cathrine Rein, Beata Grallert, Rolf Bernander, Trond Stokke, and Erik Boye. 1997. “Measurement of Nuclear DNA Content in Fission Yeast by Flow Cytometry.” *Yeast* 13: 1329–35.

Chuang, Ray Yuan, and Thomas J. Kelly. 1999. “The Fission Yeast Homologue of Orc4p Binds to Replication Origin DNA via Multiple AT-Hooks.” *Proceedings of the National Academy of Sciences of the United States of America* 96 (6): 2656–61.

Clausen, Anders R., Scott A. Lujan, Adam B. Burkholder, Clinton D. Orebaugh, Jessica S. Williams, Maryam F. Clausen, Ewa P. Malc, Piotr A. Mieczkowski, David C. Fargo, Duncan J. Smith, and Thomas A. Kunkel 2015. “Tracking Replication Enzymology in Vivo by Genome-Wide Mapping of Ribonucleotide Incorporation.” *Nature Structural and Molecular Biology* 22 (3): 185–91.

Czajkowsky, Daniel M., Jie Liu, Joyce L. Hamlin, and Zhifeng Shao. 2008. “DNA Combing

Reveals Intrinsic Temporal Disorder in the Replication of Yeast Chromosome VI." *Journal of Molecular Biology* 375 (1): 12–19.

Daigaku, Yasukazu, Thomas J. Etheridge, Yuka Nakazawa, Mayumi Nakayama, Adam T. Watson, Izumi Miyabe, Tomoo Ogi, Mark A. Osborne, and Antony M. Carr. 2017. "PCNA Ubiquitylation Ensures Timely Completion of Unperturbed DNA Replication in Fission Yeast." *PLoS Genetics* 13 (5): e1006789.

Daigaku, Yasukazu, Andrea Keszthelyi, Carolin A. Müller, Izumi Miyabe, Tony Brooks, Renata Retkute, Mike Hubank, Conrad A. Nieduszynski, and Antony M. Carr. 2015. "A Global Profile of Replicative Polymerase Usage." *Nature Structural and Molecular Biology* 22 (3): 192–98.

De Carli, Francesco, Vincent Gaggioli, Gaël A. Millot, and Olivier Hyrien. 2016. "Single-Molecule, Antibody-Free Fluorescent Visualisation of Replication Tracts along Barcoded DNA Molecules." *International Journal of Developmental Biology* 60 (7–9): 297–304.

De Carli, Francesco, Nikita Menezes, Wahiba Berrabah, Valérie Barbe, Auguste Genovesio, and Olivier Hyrien. 2018. "High-Throughput Optical Mapping of Replicating DNA." *Small Methods* 2 (9): 1800146.

Dewar, James M., and Johannes C. Walter. 2017. "Mechanisms of DNA Replication Termination." *Nature Reviews Molecular Cell Biology* 18 (8): 507–16.

Diffley, John F.X., and Julie H Cocker. 1992. "Protein-DNA Interactions at a Yeast Replication Origin." *Nature* 357 (6374): 169–72.

Diffley, John F X, Julie H Cocker, Adele Rowley, and J Dowell. 1994. "Two Steps in the Assembly of Complexes at Yeast Replication Origins In Vivo." *Cell* 78: 303–16.

Dileep, Vishnu, and David M. Gilbert. 2018. "Single-Cell Replication Profiling to Measure Stochastic Variation in Mammalian Replication Timing." *Nature Communications* 9 (1): 427.

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

Donovan, Shane, Janet Harwood, Lucy S. Drury, and John F.X. Diffley. 1997. "Cdc6p-

Dependent Loading of Mcm Proteins onto Pre-Replicative Chromatin in Budding Yeast." *Proceedings of the National Academy of Sciences of the United States of America* 94 (11): 5611–16.

Douglas, Max E., Ferdos Abid Ali, Alessandro Costa, and John F.X. Diffley. 2018. "The Mechanism of Eukaryotic CMG Helicase Activation." *Nature* 555 (7695): 265–68.

Dovrat, Daniel, Danielle Dahan, Shachar Sherman, Ioannis Tsirkas, Natalie Elia, and Amir Aharoni. 2018. "A Live-Cell Imaging Approach for Measuring DNA Replication Rates." *Cell Reports* 24 (1): 252–58.

Duina, Andrea A., Mary E. Miller, and Jill B. Keeney. 2014. "Budding Yeast for Budding Geneticists: A Primer on the Saccharomyces Cerevisiae Model System." *Genetics* 197 (1): 33–48.

Dunham, Joseph P., and Maren L. Friesen. 2013. "A Cost-Effective Method for High-Throughput Construction of Illumina Sequencing Libraries." *Cold Spring Harbor Protocols* 2013 (9): 820–34.

Ekundayo, Babatunde, and Franziska Bleichert. 2019. "Origins of DNA Replication." *PLoS Genetics* 15 (9): e1008320.

Fantes, Peter A., and Charles S. Hoffman. 2016. "A Brief History of Schizosaccharomyces Pombe Research: A Perspective over the Past 70 Years." *Genetics* 203 (2): 621–29.

Feng, Wenyi, David Collingwood, Max E. Boeck, Lindsay A. Fox, Gina M. Alvino, Walton L. Fangman, Mosur K. Raghuraman, and Bonita J. Brewer. 2006. "Genomic Mapping of Single-Stranded DNA in Hydroxyurea-Challenged Yeasts Identifies Origins of Replication." *Nature Cell Biology* 8 (2): 148–55.

Figiel, Małgorzata, Hyongi Chon, Susana M. Cerritelli, Magdalena Cybulska, Robert J. Crouch, and Marcin Nowotny. 2011. "The Structural and Biochemical Characterization of Human RNase H2 Complex Reveals the Molecular Basis for Substrate Recognition and Aicardi- Goutières Syndrome Defects." *Journal of Biological Chemistry* 286 (12): 10540–50.

Forsburg, Susan L., and Nicholas Rhind. 2006. "Basic Methods for Fission Yeast." *Yeast* 23 (3): 173–83.

Gabrieli, Tslil, Hila Sharim, Yael Michaeli, and Yuval Ebenstein. 2017. “Cas9-Assisted Targeting of CHromosome Segments (CATCH) for Targeted Nanopore Sequencing and Optical Genome Mapping.” *BioRxiv*, 110163.

Gabrieli, Tslil, Hila Sharim, Gil Nifker, Jonathan Jeffet, Tamar Shahal, Rani Arielly, Michal Levi-Sakin, Lily Hoch, Nissim Arbib, Yael Michaeli, and Yuval Ebenstein. 2018a. “Epigenetic Optical Mapping of 5-Hydroxymethylcytosine in Nanochannel Arrays.” *ACS Nano* 12 (7): 7148–58.

———. 2018b. “Genome-Wide Epigenetic Profiling of 5-Hydroxymethylcytosine by Long-Read Optical Mapping.” *BioRxiv*, 260166.

Garbacz, Marta A., Scott A. Lujan, Adam B. Burkholder, Phillip B. Cox, Qiuqin Wu, Zhi Xiong Zhou, James E. Haber, and Thomas A. Kunkel. 2018. “Evidence That DNA Polymerase δ Contributes to Initiating Leading Strand DNA Replication in Saccharomyces Cerevisiae.” *Nature Communications* 9 (1): 858.

Garg, Parie, Carrie M. Stith, Nasim Sabouri, Erik Johansson, and Peter M. Burgers. 2004. “Idling by DNA Polymerase δ Maintains a Ligatable Nick during Lagging-Strand DNA Replication.” *Genes and Development* 18 (22): 2764–73.

Ge, Xin Quan, Dean A. Jackson, and J. Julian Blow. 2007. “Dormant Origins Licensed by Excess Mcm2-7 Are Required for Human Cells to Survive Replicative Stress.” *Genes and Development* 21 (24): 3331–41.

Giaever, Guri, Angela M. Chu, Li Ni, Carla Connelly, Linda Riles, Steeve Véronneau, Sally Dow, Ankuta Lucau-Danila, Keith Anderson, Bruno André, Adam P. Arkin, Anna Astromoff, Mohamed El Bakkoury, Rhonda Bangham, Rocio Benito, Sophie Brachat, Stefano Campanaro, Matt Curtiss, Karen Davis, Adam Deutschbauer, Karl-Dieter Entian, Patrick Flaherty, Francoise Foury, David J. Garfinkel, Mark Gerstein, Deanna Gotte, Ulrich Güldener, Johannes H. Hegemann, Svenja Hempel, Zelek Herman, Daniel F. Jaramillo, Diane E. Kelly, Steven L. Kelly, Peter Kötter, Darlene LaBonte, David C. Lamb, Ning Lan, Hong Liang, Hong Liao, Lucy Liu, Chuanyun Luo, Marc Lussier, Rong Mao, Patrice Menard, Siew Loon Ooi, Jose L. Revuelta, Christopher J. Roberts, Matthias Rose, Petra Ross-Macdonald, Bart Scherens, Greg Schimmack, Brenda

Shafer, Daniel D. Shoemaker, Sharon Sookhai-Mahadeo, Reginald K. Storms, Jeffrey N. Strathern, Giorgio Valle, Marleen Voet, Guido Volckaert, Ching-yun Wang, Teresa R. Ward, Julie Wilhelmy, Elizabeth A. Winzeler, Yonghong Yang, Grace Yen, Elaine Youngman, Kexin Yu, Howard Bussey, Jef D. Boeke, Michael Snyder, Peter Philippsen, Ronald W. Davis, and Mark Johnston. 2002. "Functional Profiling of the Saccharomyces Cerevisiae Genome." *Nature*, 418 (6896): 387–91.

Gietz, R Daniel, and Robert H Schiestl. 2008. "High-Efficiency Yeast Transformation Using the LiAc / SS Carrier DNA / PEG Method." *Nature Protocols* 2 (1): 31–35.

Gilbert, David M. 2004. "In Search of the Holy Replicator." *Nature Reviews Molecular Cell Biology* 5 (10): 848–55.

Goffeau, A, B G Barrell, H Bussey, R W Davis, B Dujon, H Feldmann, F Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. 1996. "Life with 6000 Genes." *Science* 274 (5287): 546–67.

Gould, Kathleen L., and Viesturs Simanis. 1997. "The Control of Septum Formation in Fission Yeast." *Genes and Development* 11 (22): 2939–51.

Gray, Joseph V., Gregory A. Petsko, Gerald C. Johnston, Dagmar Ringe, Richard A. Singer, and Margaret Werner-Washburne. 2004. "'Sleeping Beauty': Quiescence in Saccharomyces Cerevisiae." *Microbiology and Molecular Biology Reviews* 68 (2): 187–206.

Gray, William M., J. Carlos Del Pozo, Loni Walker, Lawrence Hobbie, Eddy Risseeuw, Travis Banks, William L. Crosby, Ming Yang, Hong Ma, and Mark Estelle. 1999. "Identification of an SCF Ubiquitin-Ligase Complex Required for Auxin Response in Arabidopsis Thaliana." *Genes and Development* 13 (13): 1678–91.

Greider, Carol W. 1999. "Telomeres Do D-Loop-T-Loop." *Cell* 97 (4): 419–22.

Hagan, Iain M., Agnes Grallert, and Viesturs Simanis. 2016. "Analysis of the Schizosaccharomyces Pombe Cell Cycle." *Cold Spring Harbor Protocols*, 735–46.

Håkansson, Pelle, Lina Dahl, Olga Chilkova, Vladimir Domkin, and Lars Thelander. 2006. "The Schizosaccharomyces Pombe Replication Inhibitor Spd1 Regulates Ribonucleotide Reductase Activity and DNTPs by Binding to the Large Cdc22 Subunit."

Journal of Biological Chemistry 281 (3): 1778–83.

Hansen, R. Scott, Sean Thomas, Richard Sandstrom, Theresa K. Canfield, Robert E. Thurman, Molly Weaver, Michael O. Dorschner, Stanley M. Gartler, and John A. Stamatoyannopoulos. 2010. "Sequencing Newly Replicated DNA Reveals Widespread Plasticity in Human Replication Timing." *Proceedings of the National Academy of Sciences of the United States of America* 107 (1): 139–44.

Hartwell, Leland H., Joseph Culotti, John R. Pringle, and Brian J. Reid. 1974. "Genetic Control of the Cell Division Cycle in Yeast." *Science* 183 (4120): 46–51.

Hartwell, Leland H., and Ted A. Weinert. 1989. "Checkpoints: Controls That Ensure the Order of Cell Cycle Events." *Science* 246 (4930): 629–34.

Hawkins, Michelle, Renata Retkute, Carolin A. Müller, Nazan Saner, Tomoyuki U. Tanaka, Alessandro P.S. deMoura, and Conrad A. Nieduszynski. 2013. "High-Resolution Replication Profiles Define the Stochastic Nature of Genome Replication Initiation and Termination." *Cell Reports* 5 (4): 1132–41.

Hayashi, Makoto, Yuki Katou, Takehiko Itoh, Mitsutoshi Tazumi, Yoshiki Yamada, Tatsuro Takahashi, Takuro Nakagawa, Katsuhiko Shirahige, and Hisao Masukata. 2007. "Genome-Wide Localization of Pre-RC Sites and Identification of Replication Origins in Fission Yeast." *EMBO Journal* 26 (5): 1327–39.

Haye-Bertolozzi, Joanna E., and Oscar M. Aparicio. 2018. "Quantitative Bromodeoxyuridine Immunoprecipitation Analyzed by High-Throughput Sequencing (QBrdU-Seq or QBU)." In *Genome Instability. Methods in Molecular Biology*, edited by Marco Muzi-Falconi and G. Brown, 1672:209–25. New York: Humana Press, New York, NY.

Hayward, Daniel, Tatiana Alfonso-Pérez, and Ulrike Gruneberg. 2019. "Orchestration of the Spindle Assembly Checkpoint by CDK1-Cyclin B1." *FEBS Letters* 593 (20): 2889–2907.

Heichinger, Christian, Christopher J Penkett, Jürg Bähler, and Paul Nurse. 2006. "Genome-Wide Characterization of Fission Yeast DNA Replication Origins." *The EMBO Journal* 25 (21): 5171–79.

Hennion, Magali, Jean Michel Arbona, Laurent Lacroix, Corinne Cruaud, Bertrand Theulot, Benoît Le Tallec, Florence Proux, Xia Wu, Elizaveta Novikova, Stefan Engelen,Arnaud

Lemainque, Benjamin Audit, and Olivier Hyrien. 2020. “FORK-Seq: Replication Landscape of the Saccharomyces Cerevisiae Genome by Nanopore Sequencing.” *Genome Biology* 21: 125.

Herskowitz, Ira. 1988. “Life Cycle of the Budding Yeast Saccharomyces Cerevisiae.” *Microbiological Reviews* 52 (4): 536–53.

Hoffman, Charles S., Valerie Wood, and Peter A. Fantes. 2015. “An Ancient Yeast for Young Geneticists: A Primer on the Schizosaccharomyces Pombe Model System.” *Genetics* 201 (2): 403–23.

Hozák, Pavel, and Peter R. Cook. 1994. “Replication Factories.” *Trends in Cell Biology* 4 (2): 48–52.

Huberman, Joel A., Jiguang Zhu, Leslie R. Davis, and Carol S. Newlon. 1988. “Close Association of a DNA Replication Origin and an ARS Element on Chromosome III of the Yeast, Saccharomyces Cerevisiae.” *Nucleic Acids Research* 16 (14): 6373–84.

Huberman, Joel A, and Arthur D Riggs. 1968. “On the Mechanism of DNA Replication in Mammalian Chromosomes.” *Journal of Molecular Biology* 32: 327–41.

Hyrien, Olivier. 2015. “Peaks Cloaked in the Mist: The Landscape of Mammalian Replication Origins.” *Journal of Cell Biology* 208 (2): 147–60.

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

Jain, Miten, Hugh E. Olsen, Benedict Paten, and Mark Akeson. 2016. “The Oxford Nanopore MinION: Delivery of Nanopore Sequencing to the Genomics Community.” *Genome Biology* 17 (1): 239

Jin, Yong Hwan, Rao Ayyagari, Michael A. Resnick, Dmitry A. Gordeni, and Peter M.J. Burgers. 2003. “Okazaki Fragment Maturation in Yeast: II. Cooperation between the Polymerase and 3′-5′-Exonuclease Activities of Pol δ in the Creation of a Ligatable Nick.” *Journal of Biological Chemistry* 278 (3): 1626–33.

Johnson, Robert E., Roland Klassen, Louise Prakash, and Satya Prakash. 2015. “A Major

Role of DNA Polymerase δ in Replication of Both the Leading and Lagging DNA Strands." *Molecular Cell* 59 (2): 163–75.

———. 2016. "Response to Burgers et Al." *Molecular Cell* 61 (4): 494–95.

Johnston, Leland H., and Kim A. Nasmyth. 1978. "Saccharomyces Cerevisiae Cell Cycle Mutant Cdc9 Is Defective in DNA Ligase." *Nature* 274 (5674): 891–93.

Joyce, C. M. 1997. "Choosing the Right Sugar: How Polymerases Select a Nucleotide Substrate." *Proceedings of the National Academy of Sciences of the United States of America* 94 (5): 1619–22.

Kahli, Malik, Joseph S. Osmundson, Rani Yeung, and Duncan J. Smith. 2019. "Processing of Eukaryotic Okazaki Fragments by Redundant Nucleases Can Be Uncoupled from Ongoing DNA Replication in Vivo." *Nucleic Acids Research* 47 (4): 1814–22.

Kao, Hui I., Janaki Veeraraghavan, Piotr Polaczek, Judith L. Campbell, and Robert A. Bambara. 2004. "On the Roles of Saccharomyces Cerevisiae Dna2p and Flap Endonuclease 1 in Okazaki Fragment Processing." *Journal of Biological Chemistry* 279 (15): 15014–24.

Kaykov, Atanas, and Paul Nurse. 2015. "The Spatial and Temporal Organization of Origin Firing during the S-Phase of Fission Yeast." *Genome Research* 25 (3): 391–401.

Kaykov, Atanas, Thibaud Taillefumier, Aaron Bensimon, and Paul Nurse. 2016. "Molecular Combing of Single DNA Molecules on the 10 Megabase Scale." *Scientific Reports* 6: 19636.

Kearsey, Stephen. 1983. "Analysis of Sequences Conferring Autonomous Replication in Baker's Yeast." *The EMBO Journal* 2 (9): 1571–75.

———. 1984. "Structural Requirements for the Function of a Yeast Chromosomal Replicator." *Cell* 37 (1): 299–307.

Kim, Dong Uk, Jacqueline Hayles, Dongsup Kim, Valerie Wood, Han-Oh Park, Misun Won, Hyang-Sook Yoo, Trevor Duhig, Miyoung Nam, Georgia Palmer, Sangjo Han, Linda Jeffery, Seung-Tae Baek, Hyemi Lee, Young Sam Shim, Minho Lee, Lila Kim, Kyung-Sun Heo, Eun Joo Noh, Ah-Reum Lee, Young-Joo Jang, Kyung-Sook Chung, Shin-Jung Choi, Jo-Young Park, Youngwoo Park, Hwan Mook Kim, Song-Kyu Park, Hae-

Joon Park, Eun-Jung Kang, Hyong Bai Kim, Hyun-Sam Kang, Hee-Moon Park, Kyunghoon Kim, Kiwon Song, Kyung Bin Song, Paul Nurse, and Kwang-Lae Hoe. 2010. "Analysis of a Genome-Wide Set of Gene Deletions in the Fission Yeast Schizosaccharomyces Pombe." *Nature Biotechnology* 28 (6): 617–23.

Kim, Soo Mi, and Joel A. Huberman. 2001. "Regulation of Replication Timing in Fission Yeast." *EMBO Journal* 20 (21): 6115–26.

Kitamura, Etsushi, J. Julian Blow, and Tomoyuki U. Tanaka. 2006. "Live-Cell Imaging Reveals Replication of Individual Replicons in Eukaryotic Replication Factories." *Cell* 125 (7): 1297–1308.

Klein, Kyle, Weitao Wang, Tyler Borrman, Saki Chan, Denghong Zhang, Zhiping Weng, Alex Hastie, Chunlong Chen, David Gilbert, and Nicholas Rhind. 2017. "Genome-Wide Identification of Early-Firing Human Replication Origins by Optical Replication Mapping." *BioRxiv*, 214841.

Koh, Kyung Duk, Sathya Balachander, Jay R Hesselberth, and Francesca Storici. 2015. "Ribose-Seq: Global Mapping of Ribonucleotides Embedded in Genomic DNA." *Nature Methods* 12 (3): 251–57.

Kubota, Takashi, Kohei Nishimura, Masato T. Kanemaki, and Anne D. Donaldson. 2013. "The Elg1 Replication Factor C-like Complex Functions in PCNA Unloading during DNA Replication." *Molecular Cell* 50 (2): 273–80.

Kuchta, Robert D., and Gudrun Stengel. 2010. "Mechanism and Evolution of DNA Primases." *Biochimica et Biophysica Acta - Proteins and Proteomics* 1804 (5): 1180–89.

Kushnirov, Vitaly V. 2000. "Rapid and Reliable Protein Extraction from Yeast." *Yeast* 16 (9): 857–60.

Lacroix, Joris, Sandrine Pélofy, Charline Blatché, Marie-Jeanne Pillaire, Sébastien Huet, Catherine Chapuis, Jean-Sébastien Hoffmann, and Aurélien Bancaud. 2016. "Analysis of DNA Replication by Optical Mapping in Nanochannels." *Small* 12 (43): 5963–70.

Lam, Ernest T., Alex Hastie, Chin Lin, Dean Ehrlich, Somes K. Das, Michael D. Austin, Paru Deshpande, Han Cao, Niranjan Nagarajan, Ming Xiao, and Pui-Yan Kwok. 2012. "Genome Mapping on Nanochannel Arrays for Structural Variation Analysis and

Sequence Assembly." *Nature Biotechnology* 30 (8): 771–76.

Laskey, Ronald A., Micaela P. Fairman, and J. Julian Blow. 1989. "S Phase of the Cell Cycle." *Science* 246 (4930): 609–14.

Lebofsky, Ronald, and Aaron Bensimon. 2003. "Single DNA Molecule Analysis: Applications of Molecular Combing." *Briefings in Functional Genomics and Proteomics* 1 (4): 385–96.

Lee, J K, K Y Moon, Y Jiang, and J Hurwitz. 2001. "The Schizosaccharomyces Pombe Origin Recognition Complex Interacts with Multiple AT-Rich Regions of the Replication Origin DNA by Means of the AT-Hook Domains of the SpOrc4 Protein." *Proceedings of the National Academy of Sciences of the United States of America* 98 (24): 13589–94.

Lee, Michael E., William C. DeLoache, Bernardo Cervantes, and John E. Dueber. 2015. "A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly." *ACS Synthetic Biology* 4 (9): 975–86.

Li, Pao Chen, Louise Chretien, Jocelyn Côté, Thomas J. Kelly, and Susan L. Forsburg. 2011. "S. Pombe Replication Protein Cdc18 (Cdc6) Interacts with Swi6 (HP1) Heterochromatin Protein." *Cell Cycle* 10 (2): 323–36.

Li, Pao Chen, Marc D. Green, and Susan L. Forsburg. 2013. "Mutations Disrupting Histone Methylation Have Different Effects on Replication Timing in S. Pombe Centromere." *PLoS ONE* 8 (5): e61464.

Li, Yingfu, and Ronald R. Breaker. 1999. "Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2'-Hydroxyl Group." *Journal of the American Chemical Society* 121 (23): 5364–72.

Limas, Juanita C., and Jeanette Gowen Cook. 2019. "Preparation for DNA Replication: The Key to a Successful S Phase." *FEBS Letters* 593 (20): 2853–67.

Liu, Bochao, Jiazhi Hu, Jingna Wang, and Daochun Kong. 2017. "Direct Visualization of RNA-DNA Primer Removal from Okazaki Fragments Provides Support for Flap Cleavage and Exonucleolytic Pathways in Eukaryotic Cells." *Journal of Biological Chemistry* 292 (12): 4777–88.

Liu, Jianing, Natalie J. Nannas, Fang Fang Fu, Jinghua Shi, Brooke Aspinwall, Wayne A. Parrott, and R. Kelly Dawe. 2019. "Genome-Scale Sequence Disruption Following

Biolistic Transformation in Rice and Maize." *Plant Cell* 31 (2): 368–83.

Lock, Antonia, Kim Rutherford, Midori A. Harris, Jacqueline Hayles, Stephen G. Oliver, Jürg Bähler, and Valerie Wood. 2019. "PomBase 2018: User-Driven Reimplementation of the Fission Yeast Database Provides Rapid and Intuitive Access to Diverse, Interconnected Information." *Nucleic Acids Research* 47 (D1): D821–27.

Logsdon, Glennis A., Mitchell R. Vollger, and Evan E. Eichler. 2020. "Long-Read Human Genome Sequencing and Its Applications." *Nature Reviews Genetics* 21 (10): 597–614.

Loman, Nicholas J., and Mick Watson. 2015. "Successful Test Launch for Nanopore Sequencing." *Nature Methods* 12 (4): 303–4.

Luche, Douglas D, and Susan L Forsburg. 2009. "Cell-Cycle Synchrony for Analysis of S. Pombe." In *DNA Replication. Methods in Molecular Biology (Methods and Protocols)*, edited by S. Vengrova and J. Dalgaard, 521:437–48. Humana Press.

Lujan, Scott A., Anders R. Clausen, Alan B. Clark, Heather K. MacAlpine, David M. MacAlpine, Ewa P. Malc, Piotr A. Mieczkowski, Adam B. Burkholder, David C. Fargo, Dmitry A. Gordenin, and Thomas A. Kunkel. 2014. "Heterogeneous Polymerase Fidelity and Mismatch Repair Bias Genome Variation and Composition." *Genome Research* 24 (11): 1751–64.

Lujan, Scott A., Jessica S. Williams, Anders R. Clausen, Alan B. Clark, and Thomas A. Kunkel. 2013. "Ribonucleotides Are Signals for Mismatch Repair of Leading-Strand Replication Errors." *Molecular Cell* 50 (3): 437–43.

MacAlpine, David M. 2016. "ORChestrating the Human DNA Replication Program." *Proceedings of the National Academy of Sciences of the United States of America* 113 (33): 9136–38.

Maga, Giovanni, Giuseppe Villani, Vanessa Tillement, Manuel Stucki, Giada A. Locatelli, Isabelle Frouin, Silvio Spadari, and Ulrich Hübscher. 2001. "Okazaki Fragment Processing: Modulation of the Strand Displacement Activity of DNA Polymerase δ by the Concerted Action of Replication Protein A, Proliferating Cell Nuclear Antigen, and Flap Endonuclease-1." *Proceedings of the National Academy of Sciences of the United States of America* 98 (25): 14298–303.

Marchal, Claire, Jiao Sima, and David M. Gilbert. 2019. “Control of DNA Replication Timing in the 3D Genome.” *Nature Reviews Molecular Cell Biology* 20 (12): 721–37.

Maric, Marija, Timurs Maculins, Giacomo De Piccoli, and Karim Labib. 2014. “Cdc48 and a Ubiquitin Ligase Drive Disassembly of the CMG Helicase at the End of DNA Replication.” *Science* 346 (6208): 1253596.

Maundrell, K., A. Hutchison, and S. Shall. 1988. “Sequence Analysis of ARS Elements in Fission Yeast.” *The EMBO Journal* 7 (7): 2203–9.

Maundrell, Kinsey. 1990. “Nmtl of Fission Yeast.” *Journal of Biological Chemistry* 265 (19): 10857–64.

McCaffrey, Jennifer, Eleanor Young, Katy Lassahn, Justin Sibert, Steven Pastor, Harold Riethman, and Ming Xiao. 2017. “High-Throughput Single-Molecule Telomere Characterization.” *Genome Research* 27: 1904–15.

McGuffee, Sean R., Duncan J. Smith, and Iestyn Whitehouse. 2013. “Quantitative, Genome-Wide Analysis of Eukaryotic Replication Initiation and Termination.” *Molecular Cell* 50 (1): 123–35.

Mcintosh, J Richard, and Michael P Koonce. 1989. “Mitosis.” *Science* 246 (4930): 622–28.

Meiron, Hagit, Edna Nahon, and Dina Raveh. 1995. “Identification of the Heterothallic Mutation in HO-Endonuclease of S. Cerevisiae Using HO/Ho Chimeric Genes.” *Current Genetics* 28 (4): 367–73.

Meselson, Matthew, and Franklin W. Stahl. 1958. “The Replication of DNA in Escherichia coli.” *Proceedings of the National Academy of Sciences of the United States of America* 44 (7): 671–82.

Mitra, Sankar, and Arthur Kornberg. 1966. “Enzymatic Mechanisms of DNA Replication.” *Journal of General Physiology* 49 (6): 59–79.

Miyabe, Izumi, Thomas A. Kunkel, and Antony M. Carr. 2011. “The Major Roles of DNA Polymerases Epsilon and Delta at the Eukaryotic Replication Fork Are Evolutionarily Conserved.” *PLoS Genetics* 7 (12): e1002407.

Moreno, Sara Priego, and Agnieszka Gambus. 2020. “Mechanisms of Eukaryotic Replisome Disassembly.” *Biochemical Society Transactions* 48 (3): 823–36.

Müller, Carolin A., Michael A. Boemo, Paolo Spingardi, Benedikt M. Kessler, Skirmantas Kriaucionis, Jared T. Simpson, and Conrad A. Nieduszynski. 2019. “Capturing the Dynamics of Genome Replication on Individual Ultra-Long Nanopore Sequence Reads.” *Nature Methods* 16 (5): 429–36.

Müller, Carolin A., Michelle Hawkins, Renata Retkute, Sunir Malla, Ray Wilson, Martin J. Blythe, Ryuichiro Nakato, Makiko Komata, Katsuhiko Shirahige, Alessandro P.S. de Moura and Conrad A. Nieduszynski. 2014. “The Dynamics of Genome Replication Using Deep Sequencing.” *Nucleic Acids Research* 42 (1): e3.

Müller, Carolin A., and Conrad A. Nieduszynski. 2012. “Conservation of Replication Timing Reveals Global and Local Regulation of Replication Origin Activity.” *Genome Research* 22: 1953–62.

———. 2017. “DNA Replication Timing Influences Gene Expression Level.” *Journal of Cell Biology* 216 (7): 1907–14.

Muramatsu, Sachiko, Kazuyuki Hirai, Yon Soo Tak, Yoichiro Kamimura, and Hiroyuki Araki. 2010. “CDK-Dependent Complex Formation between Replication Proteins Dpb11, Sld2, Pol ε, and GINS in Budding Yeast.” *Genes and Development* 24 (6): 602–12.

Murray, Andrew. 1994. “Cell Cycle Checkpoints.” *Current Opinion in Cell Biology* 6 (6): 872–76.

Murray, Johanne M., Adam T. Watson, and Antony M. Carr. 2016. “Extraction of Chromosomal DNA from Schizosaccharomyces Pombe.” *Cold Spring Harbor Protocols* 2016 (5): 484–87.

Nasmyth, Kim. A. 1982. “Molecular Genetics of Yeast Mating Type.” *Annual Review of Genetics* 16: 439–500.

Nasmyth, Kim A., and David Shore. 1987. “Transcriptional Regulation in the Yeast Life Cycle.” *Science* 237 (4819): 1162–70.

Neely, Robert K., Jochem Deen, and Johan Hofkens. 2011. “Optical Mapping of DNA: Single-Molecule-Based Methods for Mapping Genomes.” *Biopolymers* 95 (5): 298–311.

Neiman, Aaron M. 2011. “Sporulation in the Budding Yeast Saccharomyces Cerevisiae.” *Genetics* 189 (3): 737–65.

Newman, Timothy J., Mohammed A. Mamun, Conrad A. Nieduszynski, and J. Julian Blow. 2013. “Replisome Stall Events Have Shaped the Distribution of Replication Origins in the Genomes of Yeasts.” *Nucleic Acids Research* 41 (21): 9705–18.

Nick McElhinny, Stephanie A., Dmitry A. Gordenin, Carrie M. Stith, Peter M.J. Burgers, and Thomas A. Kunkel. 2008. “Division of Labor at the Eukaryotic Replication Fork.” *Molecular Cell* 30 (2): 137–44.

Nick McElhinny, Stephanie A., Brian E. Watts, Dinesh Kumar, Danielle L. Watt, Else Britt Lundström, Peter M.J. Burgers, Erik Johansson, Andrei Chabes, and Thomas A. Kunkel. 2010. “Abundant Ribonucleotide Incorporation into DNA by Yeast Replicative Polymerases.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (11): 4949–54.

Nick McElhinny, Stephanie A.Nick, Dinesh Kumar, Alan B. Clark, Danielle L. Watt, Brian E. Watts, Else Britt Lundström, Erik Johansson, Andrei Chabes, and Thomas A. Kunkel. 2010. “Genome Instability Due to Ribonucleotide Incorporation into DNA.” *Nature Chemical Biology* 6 (10): 774–81.

Nieduszynski, Conrad A., Shin Ichiro Hiraga, Prashanth Ak, Craig J. Benham, and Anne D. Donaldson. 2007. “OriDB: A DNA Replication Origin Database.” *Nucleic Acids Research* 35 (Database issue): D40–46.

Nieduszynski, Conrad A., Yvonne Knox, and Anne D. Donaldson. 2006. “Genome-Wide Identification of Replication Origins in Yeast by Comparative Genomics.” *Genes and Development* 20 (14): 1874–79.

Nielsen, Olaf. 2016. “Synchronization of s Phase in Schizosaccharomyces Pombe Cells by Transient Exposure to M-Factor Pheromone.” *Cold Spring Harbor Protocols*, 768–72.

Nishimura, Kohei, Tatsuo Fukagawa, Haruhiko Takisawa, Tatsuo Kakimoto, and Masato Kanemaki. 2009. “An Auxin-Based Degron System for the Rapid Depletion of Proteins in Nonplant Cells.” *Nature Methods* 6 (12): 917–22.

Nishimura, Kohei, and Masato T. Kanemaki. 2014. “Rapid Depletion of Budding Yeast Proteins via the Fusion of an Auxin-Inducible Degron (AID).” *Current Protocols in Cell Biology* 64: 20.9.1-16.

Nurse, Paul. 2000. “A Long Twentieth Century of the Cell Cycle and Beyond.” *Cell* 100 (1): 71–78.

Nurse, Paul, Pierre Thuriaux, and Kim Nasmyth. 1976. “Genetic Control of the Cell Division Cycle in the Fission Yeast Schizosaccharomyces Pombe.” *Molecular and General Genetics* 146 (2): 167–78.

O’Donnell, Michael, Lance Langston, and Bruce Stillman. 2013. “Principles and Concepts of DNA Replication in Bacteria, Archaea, and Eukarya.” *Cold Spring Harbor Perspectives in Biology* 5 (7): a010108.

Okazaki, Reiji, Tuneko Okazaki, Kiwako Sakabe, Kazunori Sugimoto, and Akio Sugino. 1968. “Mechanism of DNA Chain Growth, I. Possible Discontinuity and Unusual Secondary Structure of Newly Synthesized Chains.” *Proceedings of the National Academy of Sciences of the United States of America* 59 (2): 598–605.

Okuno, Yukiko, Hiroyasu Satoh, Mariko Sekiguchi, and Hisao Masukata. 1999. “Clustered Adenine/Thymine Stretches Are Essential for Function of a Fission Yeast Replication Origin.” *Molecular and Cellular Biology* 19 (10): 6699–6709.

Pai, Chen Chun, Ignacio García, Shao Win Wang, Sue Cotterill, Stuart A. MacNeill, and Stephen E. Kearsey. 2009. “GINS Inactivation Phenotypes Reveal Two Pathways for Chromatin Association of Replicative A and DNA E Polymerases in Fission Yeast.” *Molecular Biology of the Cell* 20: 1213–22.

Pai, Chen Chun, Jasmin Schnick, Stuart A. MacNeill, and Stephen E. Kearsey. 2012. “Conditional Inactivation of Replication Proteins in Fission Yeast Using Hormone-Binding Domains.” *Methods* 57 (2): 227–33.

Pardee, A. B. 1974. “A Restriction Point for Control of Normal Animal Cell Proliferation.” *Proceedings of the National Academy of Sciences of the United States of America* 71 (4): 1286–90.

Pursell, Zachary F., Isabelle Isoz, Else Britt Lundström, Erik Johansson, and Thomas A. Kunkel. 2007. “Yeast DNA Polymerase ε Participates in Leading-Strand DNA Replication.” *Science* 317 (5834): 127–30.

Raghuraman, M. K., and Bonita J. Brewer. 2010. “Molecular Analysis of the Replication

Program in Unicellular Model Organisms." *Chromosome Research* 18 (1): 19–34.

Raghuraman, M K, Elizabeth A Winzeler, David Collingwood, Sonia Hunt, Lisa Wodicka, Andrew Conway, David J Lockhart, Ronald W Davis, Bonita J Brewer, and Walton L Fangman. 2001. "Replication Dynamics of the Yeast Genome." *Science* 294 (5540): 115–21.

Rang, Franka J., Wigard P. Kloosterman, and Jeroen de Ridder. 2018. "From Squiggle to Basepair: Computational Approaches for Improving Nanopore Sequencing Read Accuracy." *Genome Biology* 19 (90): 1–11.

Rath, Uttama, and David J. Sharp. 2011. "The Molecular Basis of Anaphase A in Animal Cells." *Chromosome Research* 19 (3): 423–32.

Reagan, M. S., C. Pittenger, W. Siede, and E. C. Friedberg. 1995. "Characterization of a Mutant Strain of Saccharomyces Cerevisiae with a Deletion of the RAD27 Gene, a Structural Homolog of the RAD2 Nucleotide Excision Repair Gene." *Journal of Bacteriology* 177 (2): 364–71.

Reijns, Martin A. M., Harriet Kemp, James Ding, Sophie Marion de Procé, Andrew P. Jackson, and Martin S. Taylor. 2015. "Lagging-Strand Replication Shapes the Mutational Landscape of the Genome." *Nature* 518 (7540): 502–6.

Reijns, Martin A.M., Doryen Bubeck, Lucien C.D. Gibson, Stephen C. Graham, George S. Baillie, E. Yvonne Jones, and Andrew P. Jackson. 2011. "The Structure of the Human RNase H2 Complex Defines Key Interaction Interfaces Relevant to Enzyme Function and Human Disease." *Journal of Biological Chemistry* 286 (12): 10530–39.

Remus, Dirk, Fabienne Beuron, Gökhan Tolun, Jack D. Griffith, Edward P. Morris, and John F.X. Diffley. 2009. "Concerted Loading of Mcm2-7 Double Hexamers around DNA during DNA Replication Origin Licensing." *Cell* 139 (4): 719–30.

Retkute, Renata, Conrad A. Nieduszynski, and Alessandro De Moura. 2012. "Mathematical Modeling of Genome Replication." *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 86 (3): 031916.

Rhind, Nicholas, and Paul Russell. 1998. "Mitotic DNA Damage and Replication Checkpoints in Yeast." *Current Opinion in Cell Biology* 10 (6): 749–58.

Rivera-Mulia, Juan Carlos, and David M. Gilbert. 2016. “Replication Timing and Transcriptional Control: Beyond Cause and Effect - Part III.” *Current Opinion in Cell Biology* 40: 168–78.

Ryba, Tyrone, Ichiro Hiratani, Junjie Lu, Mari Itoh, Michael Kulik, Jinfeng Zhang, Thomas C. Schulz, Allan J. Robins, Stephen Dalton, and David M. Gilbert. 2010. “Evolutionarily Conserved Replication Timing Profiles Predict Long-Range Chromatin Interactions and Distinguish Closely Related Cell Types.” *Genome Research* 20 (6): 761–70.

Rydberg, Bjorn, and John Game. 2002. “Excision of Misincorporated Ribonucleotides in DNA by RNase H (Type 2) and FEN-1 in Cell-Free Extracts.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (26): 16654–59.

Sakabe, Kiwako, and Reiji Okazaki. 1966. “A Unique Property of the Replicating Region of Chromosomal DNA.” *Biochimica et Biophysica Acta* 129 (3): 649–51.

Saner, Nazan, Jens Karschau, Toyoaki Natsume, Marek Gierliński, Renata Retkute, Michelle Hawkins, Conrad A. Nieduszynski, J. Julian Blow, Alessandro P.S. de Moura, and Tomoyuki U. Tanaka. 2013. “Stochastic Association of Neighboring Replicons Creates Replication Factories in Budding Yeast.” *Journal of Cell Biology* 202 (7): 1001–12.

Sekedat, Matthew D., David Fenyö, Richard S. Rogers, Alan J. Tackett, John D. Aitchison, and Brian T. Chait. 2010. “GINS Motion Reveals Replication Fork Progression Is Remarkably Uniform throughout the Yeast Genome.” *Molecular Systems Biology* 6: 353

Sharim, Hila, Assaf Grunwald, Tslil Gabrieli, Yael Michaeli, Sapir Margalit, Dmitry Torchinsky, Rani Arielly, Gil Nifker, Matyas Juhasz, Felix Gularek, Miguel Almalvez, Brandon Dufault, Sreetama Sen Chandra, Alexander Liu, Surajit Bhattacharya, Yi-Wen Chen, Eric Vilain, Kathryn R Wagner, Jonathan Pevsner, Jeff Reifenberger, Ernest T Lam, Alex R Hastie, Han Cao, Hayk Barseghyan, Elmar Weinhold, and Yuval Ebenstein. 2019. “Long-Read Single-Molecule Maps of the Functional Methylome.” *Genome Research* 29 (4): 646–56.

Sherman, Fred. 2002. “Getting Started with Yeast.” *Methods in Enzymology* 350: 3–41.

Sheu, Yi Jun, and Bruce Stillman. 2006. “Cdc7-Dbf4 Phosphorylates MCM Proteins via a Docking Site-Mediated Mechanism to Promote S Phase Progression.” *Molecular Cell* 24

(1): 101–13.

———. 2010. “The Dbf4-Cdc7 Kinase Promotes S Phase by Alleviating an Inhibitory Activity in Mcm4.” *Nature* 463 (7277): 113–17.

Siddiqui, Khalid, Kin Fan On, and John F X Diffley. 2013. “Regulating DNA Replication in Eukarya.” *Cold Spring Harbor Perspectives in Biology* 5 (9): a012930.

Simanis, Viesturs. 1995. “The Control of Septum Formation in Fission Yeast.” *Seminars in Cell Biology* 6 (2): 79–87.

Simpson, Jared T., Rachael E. Workman, P. C. Zuzarte, Matei David, L. J. Dursi, and Winston Timp. 2017. “Detecting DNA Cytosine Methylation Using Nanopore Sequencing.” *Nature Methods* 14 (4): 407–10.

Singh, Amanpreet, and Yong Jie Xu. 2016. “The Cell Killing Mechanisms of Hydroxyurea.” *Genes* 7 (99).

Siow, Cheuk C., Sian R. Nieduszynska, Carolin A. Müller, and Conrad A. Nieduszynski. 2012. “OriDB, the DNA Replication Origin Database Updated and Extended.” *Nucleic Acids Research* 40 (D1): 682–86.

Sivakumar, Sasirekha, Mary Porter-Goff, Prasanta K. Patel, Kristen Benoit, and Nicholas Rhind. 2004. “In Vivo Labeling of Fission Yeast DNA with Thymidine and Thymidine Analogs.” *Methods* 33 (3): 213–19.

Smith, Duncan J., and Iestyn Whitehouse. 2012. “Intrinsic Coupling of Lagging-Strand Synthesis to Chromatin Assembly.” *Nature* 483 (7390): 434–38.

Smith, Duncan J., Tejas Yadav, and Iestyn Whitehouse. 2015. “Detection and Sequencing of Okazaki Fragments in S. Cerevisiae.” In *DNA Replication. Methods in Molecular Biology*, edited by Sonya Vengrova and Jacob Dalgaard, 1300:141–53. New York: Humana Press, New York, NY.

Snedeker, Jonathan, Matthew Wooten, and Xin Chen. 2017. “The Inherent Asymmetry of DNA Replication.” *Annual Review of Cell and Developmental Biology* 33: 291–318.

Sparks, Justin L., Hyongi Chon, Susana M. Cerritelli, Thomas A. Kunkel, Erik Johansson, Robert J. Crouch, and Peter M. Burgers. 2012. “RNase H2-Initiated Ribonucleotide Excision Repair.” *Molecular Cell* 47 (6): 980–86.

Sriramachandran, Annie M., Giuseppe Petrosino, María Méndez-Lago, Axel J. Schäfer, Liliana S. Batista-Nascimento, Nicola Zilio, and Helle D. Ulrich. 2020. “Genome-Wide Nucleotide-Resolution Mapping of DNA Replication Patterns, Single-Strand Breaks, and Lesions by GLOE-Seq.” *Molecular Cell* 78 (5): 975-985.e7.

Stinchcomb, D. T., K. Struhl, and R. W. Davis. 1979. “Isolation and Characterisation of a Yeast Chromosomal Replicator.” *Nature* 282 (5734): 39–43.

Stodola, Joseph L., and Peter M. Burgers. 2016. “Resolving Individual Steps of Okazaki-Fragment Maturation at a Millisecond Timescale.” *Nature Structural and Molecular Biology* 23 (5): 402–8.

Stratmann, S. A., and A. M. Van Oijen. 2014. “DNA Replication at the Single-Molecule Level.” *Chemical Society Reviews* 43 (4): 1201–20.

Takahashi, Saori, Hisashi Miura, Takahiro Shibata, Koji Nagao, Katsuzumi Okumura, Masato Ogata, Chikashi Obuse, Shin ichiro Takebayashi, and Ichiro Hiratani. 2019. “Genome-Wide Stability of the DNA Replication Program in Single Mammalian Cells.” *Nature Genetics* 51 (3): 529–40.

Takisawa, Haruhiko, Satoru Mimura, and Yumiko Kubota. 2000. “Eukaryotic DNA Replication: From Pre-Replication Complex to Initiation Complex.” *Current Opinion in Cell Biology* 12 (6): 690–96.

Tanaka, Seiji, and John F.X. Diffley. 2002. “Interdependent Nuclear Accumulation of Budding Yeast Cdt1 and Mcm2-7 during G1 Phase.” *Nature Cell Biology* 4 (3): 198–207.

Tanaka, Seiji, Toshiko Umemori, Kazuyuki Hirai, Sachiko Muramatsu, Yoichiro Kamimura, and Hiroyuki Araki. 2007. “CDK-Dependent Phosphorylation of Sld2 and Sld3 Initiates DNA Replication in Budding Yeast.” *Nature* 445 (7125): 328–32.

Theis, James F., Chen Yang, Christopher B. Schaefer, and Carol S. Newlon. 1999. “DNA Sequence and Functional Analysis of Homologous ARS Elements of Saccharomyces Cerevisiae and S. Carlsbergensis.” *Genetics* 152 (3): 943–52.

Thömmes, Pia, and Ulrich Hübscher. 1990. “Eukaryotic DNA Replication: Enzymes and Proteins Acting at the Fork.” *European Journal of Biochemistry* 194 (3): 699–712.

Uhlmann, F. 2001. “Secured Cutting: Controlling Separase at the Metaphase to Anaphase

Transition." *EMBO Reports* 2 (6): 487–92.

Umek, Robert M., and David Kowalski. 1988. "The Ease of DNA Unwinding as a Determinant of Initiation at Yeast Replication Origins." *Cell* 52 (4): 559–67.

Urban, John M, Michael S Foulk, Cinzia Casella, and Susan A Gerbi. 2015. "The Hunt for Origins of DNA Replication in Multicellular Eukaryotes." *F1000prime Reports* 7 (March): 30.

Van Der Aa, Niels, Jiqiu Cheng, Ligia Mateiu, Masoud Zamani Esteki, Parveen Kumar, Eftychia Dimitriadou, Evelyne Vanneste, Yves Moreau, Joris Robert Vermeesch, and Thierry Voet. 2013. "Genome-Wide Copy Number Profiling of Single Cells in S-Phase Reveals DNA-Replication Domains." *Nucleic Acids Research* 41 (6): e66.

Van Houten, J V, and C S Newlon. 1990. "Mutational Analysis of the Consensus Sequence of a Replication Origin from Yeast Chromosome III." *Molecular and Cellular Biology* 10 (8): 3917–25.

Vernis, Laurence, Jure Piskur, and John F.X. Diffley. 2003. "Reconstitution of an Efficient Thymidine Salvage Pathway in Saccharomyces Cerevisiae." *Nucleic Acids Research* 31 (19): e120.

Viggiani, Christopher J., S. R V Knott, and Oscar M. Aparicio. 2010. "Genome-Wide Analysis of DNA Synthesis by BrdU Immunoprecipitation on Tiling Microarrays (BrdU-IP-Chip) in Saccharomyces Cerevisiae." *Cold Spring Harbor Protocols* 5: 2

Waga, Shou, and Bruce Stillman. 1998. "The DNA Replication Fork in Eukaryotic Cells." *Annual Review of Biochemistry* 67: 721–51.

Wang, Weitao, Kyle Klein, Karel Proesmans, Hongbo Yang, Claire Marchal, Xiaopeng Zhu, Tyler Borrman, Alex Hastie, Zhiping Weng, John Bechhoefer, Chun-Long Chen, David M. Gilbert, and Nicholas Rhind. 2020. "Genome-Wide Mapping of Human DNA Replication by Optical Replication Mapping Supports a Stochastic Model of Eukaryotic Replication Timing." *BioRxiv*, 2020.08.24.263459.

Weier, H. U. G. 2001. "DNA Fiber Mapping Techniques for the Assembly of High-Resolution Physical Maps." *Journal of Histochemistry and Cytochemistry* 49 (8): 939–48.

Wellinger, Raymund J., and Virginia A. Zakian. 2012. "Everything You Ever Wanted to Know

about Saccharomyces Cerevisiae Telomeres: Beginning to End." *Genetics* 191 (4): 1073–1105.

Williams, Jessica S., Scott A. Lujan, and Thomas A. Kunkel. 2016. "Processing Ribonucleotides Incorporated during Eukaryotic DNA Replication." *Nature Reviews Molecular Cell Biology* 17 (6): 350–63.

Wohlgemuth, Jay G., Gabriella H. Bulboaca, Mohamad Moghadam, Mark S. Caddie, and Michele P. Calos. 1994. "Physical Mapping of Origins of Replication in the Fission Yeast Schizosaccharomyces Pombe." *Molecular Biology of the Cell* 5 (8): 839–49.

Wong, Karen H.Y., Walfred Ma, Chun Yu Wei, Erh Chan Yeh, Wan Jia Lin, Elin H. F. Wang, Jen Ping Su, Feng-Jen Hsieh, Hsiao-Jung Kao, Hsiao-Huei Chen, Stephen K. Chow, Eleanor Young, Catherine Chu, Annie Poon, Chi-Fan Yang, Dar-Shong Lin, Yu-Feng Hu, Jer-Yuarn Wu, Ni-Chung Lee, Wuh-Liang Hwu, Dario Boffelli, David Martin, Ming Xiao, and Pui-Yan Kwok. 2020. "Towards a Reference Genome That Captures Global Genetic Diversity." *Nature Communications* 11: 5482.

Wood, V., R. Gwilliam, M.-A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, S. Baker, D. Basham, S. Bowman, K. Brooks, D. Brown, S. Brown, T. Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, P. Davis, T. Feltwell, A. Fraser, S. Gentles, A. Goble, N. Hamlin, D. Harris, J. Hidalgo, G. Hodgson, S. Holroyd, T. Hornsby, S. Howarth, E. J. Huckle, S. Hunt, K. Jagels, K. James, L. Jones, M. Jones, S. Leather, S. McDonald, J. McLean, P. Mooney, S. Moule, K. Mungall, L. Murphy, D. Niblett, C. Odell, K. Oliver, S. O'Neil, D. Pearson, M. A. Quail, E. Rabbinowitsch, K. Rutherford, S. Rutter, D. Saunders, K. Seeger, S. Sharp, J. Skelton, M. Simmonds, R. Squares, S. Squares, K. Stevens, K. Taylor, R. G. Taylor, A. Tivey, S. Walsh, T. Warren, S. Whitehead, J. Woodward, G. Volckaert, R. Aert, J. Robben, B. Grymonprez, I. Weltjens, E. Vanstreels, M. Rieger, M. Schäfer, S. Müller-Auer, C. Gabel, M. Fuchs, C. Fritzc, E. Holzer, D. Moestl, H. Hilbert, K. Borzym, I. Langer, A. Beck, H. Lehrach, R. Reinhardt, T. M. Pohl, P. Eger, W. Zimmermann, H. Wedler, R. Wambutt, B. Purnelle, A. Goffeau, E. Cadieu, S. Dréano, S. Gloux, V. Lelaure, S. Mottier, F. Galibert, S. J. Aves, Z. Xiang, C. Hunt, K. Moore, S. M. Hurst, M. Lucas, M. Rochet, C. Gaillardin, V.

A. Tallada, A. Garzon, G. Thode, R. R. Daga, L. Cruzado, J. Jimenez, M.Sánchez, F. del Rey, J. Benito, A. Domínguez, J. L. Revuelta, S. Moreno, J. Armstrong, S. L. Forsburg, L. Cerrutti, T. Lowe, W. R. McCombie, I. Paulsen, J. Potashkin, G. V. Shpakovski, D. Ussery, B. G. Barrell, and P. Nurse. 2002. “The Genome Sequence of Schizosaccharomyces Pombe.” *Nature* 415 (6874): 871–80.

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

Xu, Jia, Yoshimi Yanagisawa, Alexander M. Tsankov, Christopher Hart, Keita Aoki, Naveen Kommajosyula, Kathleen E. Steinmann, James Bochicchio, Carsten Russ, Aviv Regev, Oliver J Rando, Chad Nusbaum, Hironori Niki, Patrice Milos, Zhiping Weng, and Nicholas Rhind. 2012. “Genome-Wide Identification and Characterization of Replication Origins by Deep Sequencing.” *Genome Biology* 13 (4): R27.

Yekezare, Mona, Belén Gómez-González, and John F X Diffley. 2013. “Controlling DNA Replication Origins in Response to DNA Damage - Inhibit Globally, Activate Locally.” *Journal of Cell Science* 126 (6): 1297–1306.

Yoshida, Kazumasa, Ana Poveda, and Philippe Pasero. 2013. “Time to Be Versatile: Regulation of the Replication Timing Program in Budding Yeast.” *Journal of Molecular Biology* 425 (23): 4696–4705.

Yuan, Yuxuan, Claire Yik Lok Chung, and Ting Fung Chan. 2020. “Advances in Optical Mapping for Genomic Research.” *Computational and Structural Biotechnology Journal* 18: 2051–62.

Zegerman, Philip, and John F.X. Diffley. 2007. “Phosphorylation of Sld2 and Sld3 by Cyclin-Dependent Kinases Promotes DNA Replication in Budding Yeast.” *Nature* 445 (7125): 281–85.

Zetterberg, A., and O. Larsson. 1985. “Kinetic Analysis of Regulatory Events in G1 Leading to Proliferation or Quiescence of Swiss 3T3 Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 82 (16): 5365–69.

Zhu, Jiguang, Christine Brun, Hisanori Kurooka, Mitsuhiro Yanagida, and Joel A. Huberman. 1992. “Identification and Characterization of a Complex Chromosomal Replication Origin in Schizosaccharomyces Pombe.” *Chromosoma* 102 (1 Supplement): S7–16.