Studying DNA replication dynamics in *Schizosaccharomyces pombe* using next generation sequencing

Katarzyna (Katie) Ptasińska

Submitted for the degree of Doctor of Philosophy

University of Sussex

May 2018

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Signature: _____

Date: _____

Acknowledgements

First and foremost, I would like to thank the Medical Research Council for providing the funding that allowed me to carry out this work (grant code MR/N50189X/1).

I would like to thank Tony Carr for hiring me, not firing me and being a brilliant and supportive supervisor throughout. Thank you for letting me interrupt your days to show you pictures of lines and for not making me do Western or Southern blots.

A huge thank you, all the love and many hugs to Andrea Keszthelyi. Thank you for being so incredibly patient and supportive and for helping me remain (relatively) sane and stable. You're the best and none of this would have worked without you.

Thank you to all members of the Carr and Murray labs for the advice, support and all the reagents that that you guys let me "borrow". Thank you to all my lovely friends (who can all get equally offended for not being mentioned by name) for the wine, tears and love that got me through the days when cells weren't growing, cloning was failing and computers were crashing.

Special thanks to the family for motivating and supporting me throughout. Toluniu, thank you for being the wonderfully amazing person that you are and for spending years of your life teaching me to never divide by zero. Andrzej, you never stopped inspiring my need to question the world around me and to grow with it. Misiek, you've always been my support network and a much needed voice of reason. As a thank you, I dedicate this pinnacle of my achievements so far to you <3.

Last, but not least, thank you to Laura and Patrick. You welcomed me into your home and family, making my experience in Brighton so much more wonderful and magical than I could have ever hoped for. Thank you for all the love and prosecco and for always reminding me that there is more to life than yeast.

UNIVERSITY OF SUSSEX

Katarzyna (Katie) Ptasińska

A thesis submitted for the degree of Doctor of Philosophy

Studying DNA replication dynamics in *Schizosaccharomyces pombe* using next generation sequencing

Summary

Regions at which replication is initiated are called origins of replication and the fidelity of DNA replication is crucial to genome stability. A number of factors can affect the progression of DNA replication and origin firing. In unperturbed cells, origins are fired in a fixed temporal manner, a phenomenon known as the DNA replication timing program.

In this thesis, a novel next generation deep sequencing technique was optimised to study the progression of DNA replication through repetitive and heterochromatic regions in a variety of genetic backgrounds. The loss of Rif1, a protein implicated in a number of processes, such as facultative heterochromatin formation, DNA damage response and DNA replication timing control, yielded the most unexpected results. In addition to local effects in origin firing activity around Rif1 binding sites, *rif1*Δ resulted in a complete loss of the global replication timing program. Based on these data, this thesis further explores the relationship between the global replication timing program and the landscape of origin firing efficiencies.

In metazoans, the establishment of the DNA replication timing program was linked to the nuclear distribution of chromatin. Here, we describe the role that the tethering of chromatin to the nuclear periphery plays in establishing the global replication timing program in *S. pombe*. Finally, we present a model explaining the global replication timing program in *S. pombe* and the role that global origin firing plays within it.

Table of contents

List of figures

List of tables

Chapter 1

1.1 General introduction and thesis aims

The genetic information describing the fundamental characteristics of life, known as the genome, is stored in each cell in the form of deoxyribonucleic acid (DNA). The replication of DNA and its segregation into newly formed daughter cells is the basis of growth and reproduction.

The fidelity of DNA replication is crucial to the genomic stability of cells. A number of endogenous and exogenous agents can affect the accuracy of DNA replication. Cells have, therefore, evolved a complex series of mechanisms to limit the damage that can accrue when replication is perturbed.

Cells replicate their DNA in a fixed temporal order, a phenomenon known as DNA replication timing. The agents involved in maintaining this order and the consequences of affecting it are still elusive. Similarly to other forms of DNA replication stress, the loss of the DNA replication timing program has been linked to erroneous replication and genome instability diseases, such as cancer.

The central aim of the work for this thesis was to study the DNA replication dynamics in the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) using deep-sequencing techniques. Two main questions were explored. Firstly, do heterochromatin and repetitive DNA sequences act as barriers to DNA replication and perturb the passage of replication forks? Secondly, what are the mechanisms controlling DNA replication timing and firing of replication origins?

This chapter has been divided into three parts. Part one will introduce and discuss DNA replication and the mechanisms that are in place to limit the damage resulting from perturbed replication. Part two will focus on the formation and replication of heterochromatin. Finally, part three will discuss the DNA replication timing program, the current mathematical models describing it and the link between DNA replication timing and chromatin organisation.

The protein names in the text will follow *S. pombe* nomenclature, unless otherwise specified. References to *Saccharomyces cerevisiae* (*S. cerevisiae*) and human specific names will be made using the *sc* and *h* prefixes, respectively.

1.2 DNA replication and consequences of replication perturbation

1.2.1 Initiation of DNA replication

In 1963, Jacob and Brenner proposed the "replicon model" to explain DNA replication initiation in bacteria (Jacob and Brenner 1963). In their model, DNA replication started at *cis*-acting DNA sequences that were bound and activated by *trans*-acting factors. Studies have shown that this simple model is not only true for bacteria but is also highly conserved in all domains on life (O'Donnell, Langston et al. 2013). The complexity of the regulatory proteins and pathways involved in this process is greatest in eukaryotes.

The initiation of eukaryotic DNA replication is two-tiered. First, the *trans*-acting initiation factors recognise and bind to the *cis*-acting DNA replication origins. This step, called 'origin licensing', happens in the G_1 phase of the cell cycle and marks the loci that can act as DNA replication origins. At the onset of S-phase, a number of the licensed origins are activated and "fire" to initiate DNA replication.

1.2.1.1 Origin licensing

Eukaryotic origins have been shown to correlate with a number of factors but are generally not defined by a single DNA sequence (O'Donnell, Langston et al. 2013). One notable exception is the budding yeast *S. cerevisiae*, whose *cis*-acting autonomously replicating sequence (ARS) confers full origin activity, i.e., is capable of initiating replication when cloned into plasmids (Dhar, Sehgal et al. 2012). In *S. pombe*, origins tend to correlate with low transcriptional activity, high A-T richness and open chromatin (Dai, Chuang et al. 2005).

Origins of replication are bound in G_1 by a six subunit origin recognition complex (ORC) and a related origin binding protein Cdc18 (*sc*Cdc6) (Kelly, Martin et al. 1993, Masai, Matsumoto et al. 2010). This is followed by the loading of two hexameric Mcm2-7 (minichromosome maintenance complex) helicases, each bound by Cdt1 (Masai, Matsumoto et al. 2010). The Orc6 subunit of the ORC complex interacts with Cdt1, allowing the loading of Mcm2-7 onto DNA (Chen, de Vries et al. 2007). This interaction triggers ATP hydrolysis by Cdc18 (*sc*Cdc6) (Randell, Bowers et al. 2006) and the subsequent ejection of Cdc18 (*sc*Cdc6) and Cdt1 (O'Donnell, Langston et al. 2013). A recent biochemical and cryo-EM study proposed an "acrobat" model of Mcm2-7 loading, which involves the consecutive

loading of the two Mcm2-7 hexamers and two rounds of Cdc18 (*sc*Cdc6) release (Zhai, Li et al. 2017). The resultant Mcm2-7 and ORC complex, called the pre-replicative complex (pre-RC), remains inactive until the onset of S-phase (Fig. 1.1 A) (O'Donnell, Langston et al. 2013).

The formation of the pre-RC is restricted to G_1 phase to prevent re-replication, often using a number of redundant mechanisms. Cyclin dependent kinase (CDK) activity is associated with pre-RC formation in most organisms studied to date (O'Donnell, Langston et al. 2013). In higher eukaryotes, outside G_1, Cdt1 is tightly associated with Geminin. This sequesters Cdt1 away from Mcm2-7, preventing origin licencing outside G_1 (Wohlschlegel, Dwyer et al. 2000). In *S. pombe* and metazoans, Cdt1 is marked for proteolysis in S-phase by the action of Cul4 ubiquitin ligase (Zhong, Feng et al. 2003, Nishitani, Sugimoto et al. 2006, Ralph, Boye et al. 2006). Additionally, in *S. pombe*, the expression of Cdt1 (Hofmann and Beach 1994) and Cdc18 (Baum, Nishitani et al. 1998) is limited to S-phase through the action of G_1/S transcription factor Cdc10.

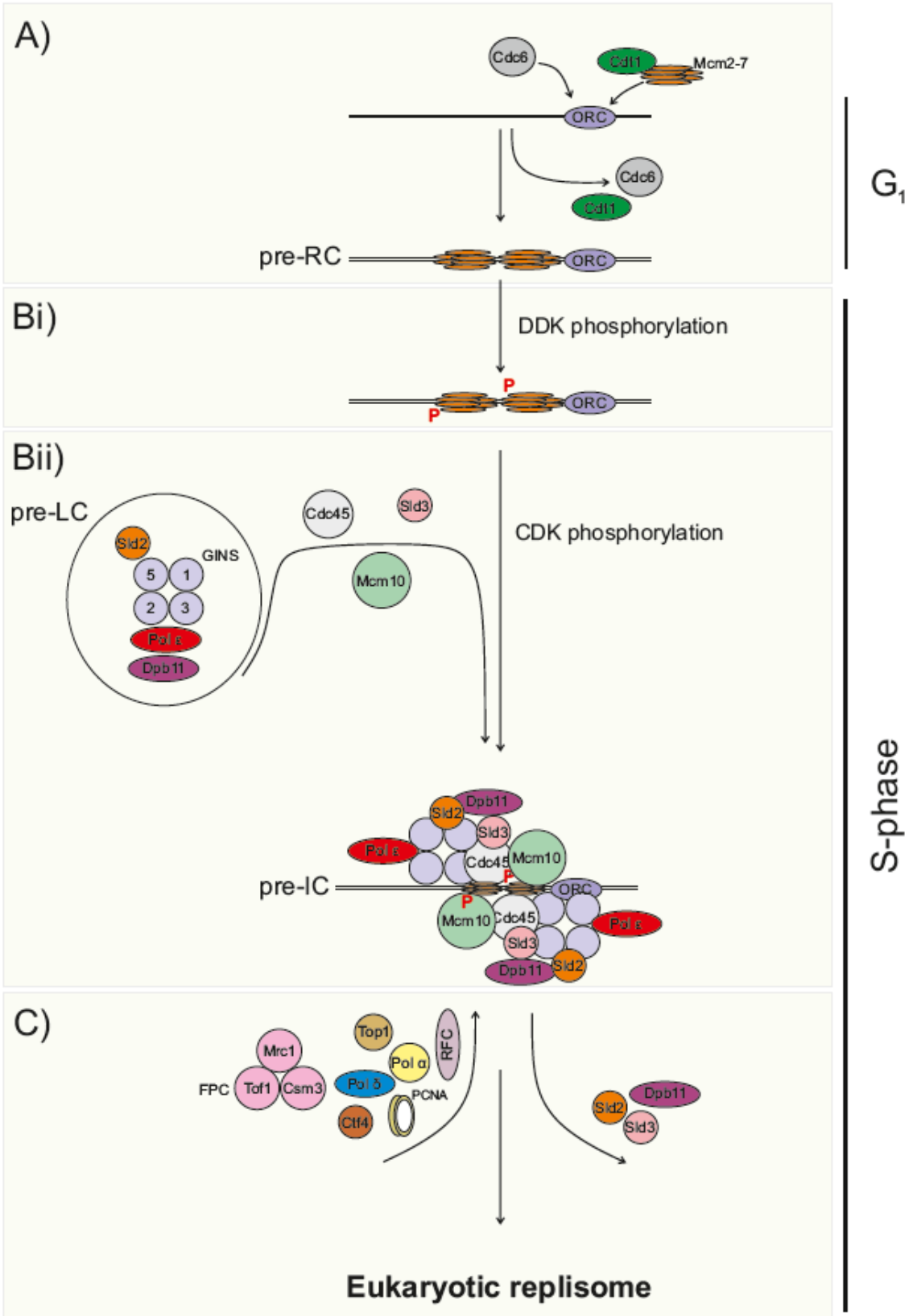

A)
Cdc6
Cdt1
Mcm2-7
ORC
Cdc6
Cdt1
pre-RC
ORC
G1
Bi)
DDK phosphorylation
P
P
ORC
Bii)
pre-LC
Sld2
GINS
5
1
2
3
Pol ε
Dpb11
Cdc45
Sld3
Mcm10
CDK phosphorylation
Dpb11
Sld2
Sld3
Pol ε
Cdc45
Mcm10
P
pre-IC
ORC
P
Mcm10
Cdc45
Pol ε
Sld3
Dpb11
Sld2
S-phase
C)
Mrc1
FPC
Tof1
Csm3
Top1
Pol α
Pol δ
PCNA
Ctf4
RFC
Sld2
Dpb11
Sld3
Eukaryotic replisome

Fig. 1.1 - **Eukaryotic (*S. cerevisiae*) DNA replication initiation pathway**

The initiation of DNA replication has been best described in *S. cerevisiae*. Although some details differ between organisms (differences are discussed in the text) the principles of the pathway are highly conserved between eukaryotes (O'Donnell, Langston et al. 2013).

A) In G_1, ORC binds origins of replication. This recruits Cdc6 and subsequently, Cdt1 bound Mcm2-7 helicase. Two Mcm2-7 helicases bind to each ORC, forming the pre-initiation complex (pre-RC).

B) The pre-RC is activated in S-phase by the DDK dependent phosphorylation of Mcm2-7. This allows the association of a number of replication initiation factors with the pre-RC, many of them binding in a CDK-dependent manner. GINS, Sld2, Dpb11 and polymerase ε are thought to associate together in a pre-loading complex (pre-LC). The pre-LC, along with Cdc45, Sld3 and Mcm10 bind the pre-RC to form the pre-initiation complex (pre-IC).

C) After pre-IC formation, Dpb11, Sld2 and Sld3 dissociate and do not travel with the replication fork. Polymerases δ and α, Tof1, Csm3, Mrc1, Ctf4, Top1, RFC and PCNA bind to form the complete eukaryotic replisome.

Adapted from (O'Donnell, Langston et al. 2013).

1.2.1.2 Origin activation

The activation of the pre-RC can be divided into two stages - the phosphorylation of the Mcm2-7 helicase and the downstream recruitment of replication initiation factors.

The DDK dependent phosphorylation of Mcm2-7 subunits has been linked to the activation of the helicase (Fig. 1.1 Bi) (Lei, Kawasaki et al. 1997, Jiang, McDonald et al. 1999, Masai, Matsui et al. 2000, Sheu and Stillman 2010). DDK activity is driven by the association of Hsk1 kinase with regulatory subunit Dfp1 (*sc*Cdc7 and *sc*Dbf4, respectively) (Larasati and Duncker 2016). The expression of Dfp1 rises at the G_1/S transition, remains high during S-phase and decreases as the cells enter M-phase (Brown and Kelly 1999), limiting the activity of DDK to S-phase.

The exact contribution of the phosphorylation of every Mcm2-7 subunit to the activation of helicase has not yet been fully elucidated. The hyper-phosphorylation of Mcm4, which is necessary but not sufficient for the onset of replication, is thought to relieve the inhibition of Mcm2-7 (Sheu and Stillman 2010). In *S. cerevisiae*, the phosphorylation of Mcm4 was shown to be stimulated by Mcm10 (Mcm10 is not part of the replicative helicase or related to Mcm2-7) (Perez-Arnaiz, Bruck et al. 2016).

After helicase activation, a number of initiation factors are recruited to form the pre-initiation complex (pre-IC) and then, after the loading of DNA polymerases and DNA melting, the full eukaryotic replisome (Fig. 1.1 B-C) (O'Donnell, Langston et al. 2013). The order in which the initiation factors bind and their interactions with each other may differ between organisms and are the subject of current research.

In *S. pombe*, the DDK dependent binding of Sld3 was proposed to initiate the pre-IC assembly. The recruitment of the GINS complex (Sld5, Psf1, Psf2, and Psf3), Rad4 (*sc*Dpb11*; h*TopBP1) and Cdc45 was shown to be dependent on Sld3 in *S. pombe* (Nakajima and Masukata 2002, Yabuuchi, Yamada et al. 2006). In *S. cerevisiae*, on the other hand, the binding of Sld3 and Cdc45 is simultaneous (Kamimura, Tak et al. 2001). In yeast and metazoans, the interaction of Cdc45 with the pre-RC was shown to be stimulated by Cdc23 (*sc*Mcm10) (Wohlschlegel, Dhar et al. 2002, Gregan, Lindner et al. 2003, Sawyer, Cheng et al. 2004, Perez-Arnaiz, Bruck et al. 2016).

The formation of the pre-IC is also partially controlled by the activity of CDK (O'Donnell, Langston et al. 2013). In the case of *S. pombe*, CDK activity is driven throughout the cell cycle by a single kinase (Cdc2). The specificity of CDK phosphorylation is controlled by the association of Cdc2 with a number of regulatory cyclins, whose expression profiles oscillate throughout the cell cycle (Coudreuse and Nurse 2010). CDK phosphorylation of Drc1 (*sc*Sld2) in *S. pombe* (Noguchi, Shanahan et al. 2002) and Sld2 and Sld3 in *S. cerevisiae* (Zegerman and Diffley 2007) was shown to control DNA replication by promoting the association of Rad4 (*sc*Dpb11) with the pre-RC. In *S. cerevisiae*, it was proposed that Dpb11, CDK phosphorylated Sld2, GINS and Polymerase ε form the so-called pre-loading complex (pre-LC) and bind the pre-RC together (Muramatsu, Hirai et al. 2010).

Cdc45, the activated Mcm2-7 helicase and the GINS complex form the 11 subunit active CMG replicative helicase, which was first characterised in the fruit fly *Drosophila melanogaster* (*D. melanogaster*) (Moyer, Lewis et al. 2006). Neither human nor *D. melanogaster* Mcm2-7 displays helicase activity in the absence of Cdc45 and GINS (Ilves, Petojevic et al. 2010, Kang, Galal et al. 2012). In *S. pombe*, the deletion of the catalytic subunit of polymerase ε (*cdc20*Δ) prevents the formation of CMG (Handa, Kanke et al. 2012), suggesting that a pre-LC-like complex may also form in fission yeast.

1.2.2 The eukaryotic replisome

Recently, a partial eukaryotic replisome was reconstituted *in vitro* using 16 purified *S. cerevisiae* proteins (Yeeles, Deegan et al. 2015). This replication system, lacked several known replisome associated factors and is, therefore, not fully descriptive of eukaryotic DNA replication. It is, however, the most comprehensive *in vitro* study of the eukaryotic stepwise replication initiation process to date. (Yeeles, Deegan et al. 2015) showed that after ORC and Mcm2-7 binding, Sld3, Sld7 and Cdc45 are recruited to the pre-RC in a DDK-dependent manner. This is followed by the binding of Sld2, Dpb11, GINS, polymerase ε and Mcm10 and requires CDK and DDK activity. This set of factors was sufficient to unwind the double helix enough to recruit the single stranded DNA (ssDNA) binding protein RPA.

It is important to note that despite being necessary for the establishment on the pre-IC, Sld2, Sld3 and Dpb11 do not travel with the replisome (Gambus, Jones et al. 2006). All three proteins have a higher affinity to ssDNA than to Mcm2-7 and upon melting of the dsDNA at the origin of replication they dissociate from the pre-IC (Bruck and Kaplan 2011, Bruck and Kaplan 2014, Dhingra, Bruck et al. 2015).

Mass spectrometry analysis of the factors associated with the *S. cerevisiae* CMG, showed that GINS, Ctf4, Tof1, Csm3, Mrc1, Top1, Mcm10, FACT and Pol α all associate with and move with the replisome (Gambus, Jones et al. 2006). This suggests that the CMG interacts with regulatory proteins, forming a large replisome progression complex (RPC) that travels to unwind the dsDNA duplex and replicate DNA. While these factors are important for the regulation of DNA replication (O'Donnell and Li 2016, Yeeles, Janska et al. 2017), the assembly of DNA polymerases at the leading and lagging strands is independent of all auxiliary factors other than PCNA (proliferating cell nuclear antigen) and the clamp loader RFC (Georgescu, Langston et al. 2014).

Eukaryotic DNA replication requires three B family DNA polymerases, all of which play different roles at the replisome (O'Donnell and Li 2016). Polymerase α has RNA primase activity (Conaway and Lehman 1982) and primes replication on both the leading and lagging strands (Georgescu, Langston et al. 2014). Polymerase ε is a highly processive polymerase, whose interaction with the CMG stabilises it on the leading strand (Georgescu, Langston et al. 2014). Polymerase δ, on the other hand, has a higher affinity for the lagging strand, when the DNA is associated with PCNA and RPA (Georgescu, Langston et al. 2014). This asymmetric assembly of the polymerases results in the division of labour at the replication fork, with polymerases ε and δ replicating the leading and lagging strands, respectively (Nick McElhinny, Gordenin et al. 2008, Miyabe, Kunkel et al. 2011, Daigaku, Keszthelyi et al. 2015).

Recent *in vitro* work suggests that polymerase δ may also plays a role in establishing leading strand synthesis before the association of the leading strand with polymerase ε (Yurieva and O'Donnell 2016, Yeeles, Janska et al. 2017). Polymerase δ activity on both the leading and lagging strands around origins of replication has been reported in *S. pombe* and *S. cerevisiae* (Daigaku, Keszthelyi et

al. 2015, Garbacz, Lujan et al. 2018). Additionally, an analysis of mutational bias in human cells suggested that polymerase δ is active on both strands ~ 1 kb around origins of replication (Artem V. Artemov, Maria A. Andrianova et al. 2017). At the time of writing this thesis, however, this work has not yet been peer-reviewed.

1.2.3 Replication fork stalling and collapse

At each origin of replication, two RFCs move bi-directionally away from each other, unwinding the dsDNA ahead of DNA polymerases. This generates two replication forks, each with continuous replication on the leading strand and discontinuous replication on the lagging strand - shown in Fig. 1.2.

The progression of replication forks can be stalled by various obstacles, collectively referred to here as replication fork barriers (RFBs). The cellular response to impeded fork progression depends on the nature of the RFB (Lambert and Carr 2013). RFBs can be intrinsic (i.e., caused by an endogenous chromosomal features) and exogenous (e.g., caused by DNA damaging agents).

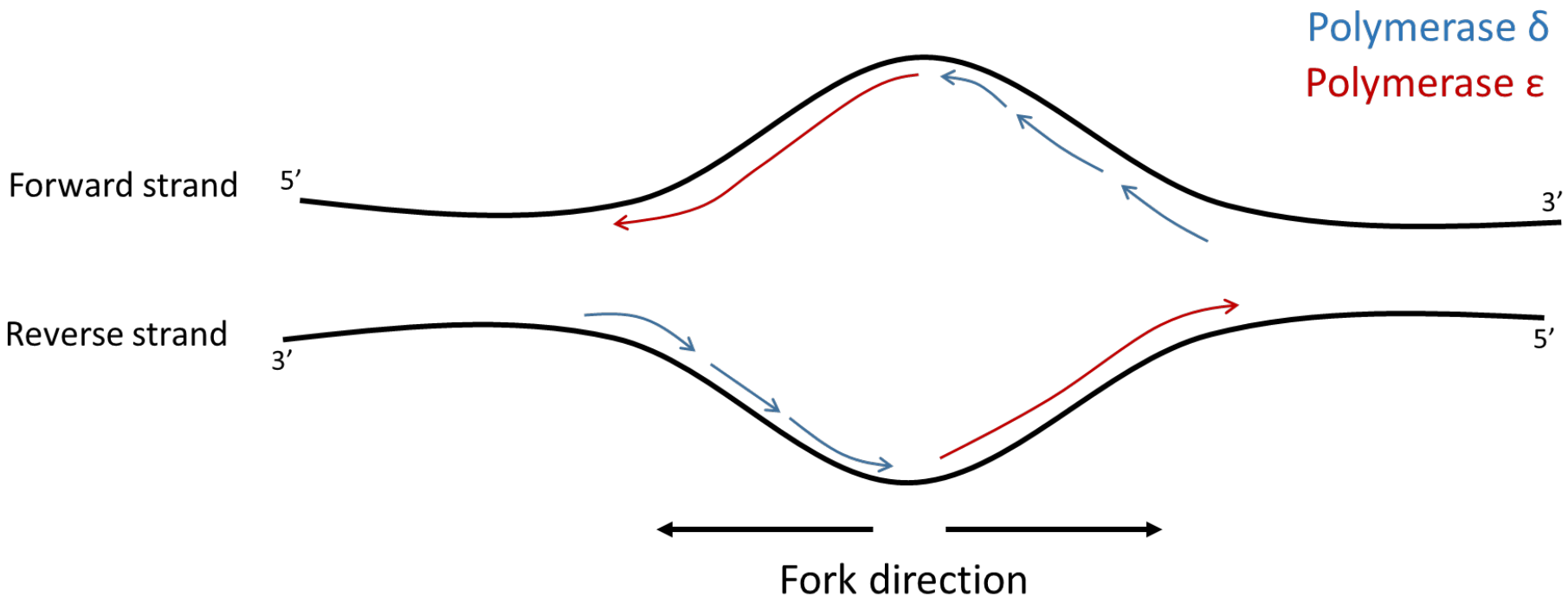


Fig. 1.2 – **Canonical DNA replication bubble**

Two replication forks travel in opposite directions from a single origin of replication. Due to the fact that DNA can only be polymerised in a 5’ to 3’ fashion, each replication fork has a lagging and leading strand, replicated by polymerases δ and ε, respectively. The newly synthesised strands polymerised by polymerases δ and ε are shown in blue and red, respectively.

1.2.3.1 Replication fork stalling at intrinsic RFBs

Both *S. pombe* and *S. cerevisiae* have evolved programmed RFBs that can stall the progression of replication forks. The evolutionarily conserved fork protection complex (FPC) is critical to this process. The FPC, composed of Swi1 and Swi3 (*sc*Tof1 and *sc*Csm3, respectively; *h*Timeless and *h*Tipin, respectively) travels with the replication fork (Noguchi, Noguchi et al. 2003, Noguchi, Noguchi et al. 2004) and directly interacts with Mrc1 (Shimmoto, Matsumoto et al. 2009).

In *S. cerevisiae*, a polar rDNA barrier (bound by Fob1) prevents head on collisions between replication forks and the transcriptional machinery by pausing fork progression in one direction (Kobayashi and Horiuchi 1996). Stalling at this programmed RFB does not lead to the disassembly of the replisome, allowing replication to restart in an HR-independent fashion (Calzada, Hodgson et al. 2005). Pausing at the rDNA barrier is dependent on Tof1 and Csm3 but not Mrc1 (Calzada, Hodgson et al. 2005, Tourriere, Versini et al. 2005). Both Tof1 and Csm3 prevent the displacement of Fob1, thereby ensuring fork pausing at the barrier (Mohanty, Bairwa et al. 2006).

In *S. pombe*, Swi1 and Swi3 are involved in stalling replication forks at the intergenic spacer regions of rDNA (Krings and Bastia 2004) and at the *RTS1* (replication termination site 1) barrier. The binding of Rtf1 to the *RTS1* barrier allows the FPC dependent stalling of forks at the mating type (*mat*) locus, which is necessary for mating type switching (Dalgaard and Klar 2001, Eydmann, Sommariva et al. 2008).

The FPC has also been linked to the maintenance of genome stability at repetitive regions. In this case, however, instead of allowing stable fork stalling, the FPC promotes fork progression. Replication fork stalling at *Alu*, Friedreich's ataxia GAA repeats and CGG repeats inserted into *S. cerevisae* genome increased in the absence of the FPC (Voineagu, Narayanan et al. 2008, Shishkin, Voineagu et al. 2009, Voineagu, Surka et al. 2009). It also, however, led to an increase in the deleterious expansion of the repeats, suggesting that FPC activity at repeats is involved in controlling their expansion. Similarly, the depletion of Timeless and Tipin led to an increase in the expansion of repeats associated with myotonic dystrophy type 1 in human cells (Liu, Chen et al. 2012).

1.2.3.2 Replication fork stalling in response to DNA replication stress

In most cases, replication stress results in transient pausing. Lesions formed on the lagging strand can be bypassed easily by re-priming replication using a new Okazaki fragment downstream of the lesion. Despite leading strand replication being mostly continuous, in some cases replication can be re-primed downstream of a lesion on the leading strand as well (reviewed in (Guilliam and Doherty 2017)). Alternatively, replication can be rescued by converging replication forks (Lambert and Carr 2013). When these primary rescue mechanisms fail, the intra-S-phase checkpoint is activated, inhibiting the firing of late origins and stalling DNA replication (Lambert and Carr 2005).

1.2.3.2.1 The intra-S-phase checkpoint

When RFBs inhibit polymerases (or polymerisation is downregulated, e.g., by depleting the deoxyribonucleotide (dNTP) pool by hydroxyurea (HU)), the helicase can carry on unwinding the duplex. This generates ssDNA, which subsequently associates with RPA (Lambert and Carr 2005). In yeast and humans, the ssDNA-RPA complex is the signal for PI3-kinase Rad3 (*sc*Mec1, *h*ATR), which associates with Rad26 (*sc*Lcd1; *hATRIP*) (Edwards, Bentley et al. 1999), to phosphorylate downstream substrates and propagate the intra-S-phase checkpoint signal (Bakkenist and Kastan 2004). The signal is ultimately relayed to effector kinase Cds1 (*sc*Rad53; *h*CHK1) via Mrc1 (*h*Claspin) (Alcasabas, Osborn et al. 2001, Tanaka and Russell 2001, Chini and Chen 2003).

Collapsed replication forks (in this thesis, a "collapsed replication fork" will be used to describe a fork that is no longer associated with DNA synthesis) can be processed and lead to aberrant DNA structures (Carr, Paek et al. 2011). These structures have been traditionally visualised using 2D gel electrophoresis. An increase in stalled replication forks and aberrant DNA structures in the absence of the intra-S-phase checkpoint has been extensively documented in *S. pombe* and *S. cerevisiae* (Carr, Paek et al. 2011). These data led to the hypothesis that intra-S-phase checkpoint played a direct role in the prevention of RFC disassembly, thereby preventing fork collapse (Zegerman and Diffley 2009). Current data, however, suggest that this may not be the case.

A comprehensive ChIP-Seq study of replisome stability in *S. cerevisiae* (De Piccoli, Katou et al. 2012) showed that the PI-3 kinase Mec1 (Chatre, Fernandes et al.) and the effector kinase Rad53 are not responsible for stabilising replisomes that stall in HU. Instead, the data suggest that the intra-S-phase checkpoint prevents the sliding of the stalled replisome along the DNA at early firing origins (De Piccoli, Katou et al. 2012). Consistently with these data, a quantitative mass spectrometry approach showed that replisome stability is not directly regulated by ATR in human cells (Dungrawala, Rose et al. 2015).

1.2.3.2.2 The role of the FPC after checkpoint activation

In *S. cerevisiae*, the aberrant structures (measured using 2D gel electrophoresis) that form in response to HU exposure, do not increase to the same extend in *mrc1*Δ and *tof1*Δ cells, compared to a checkpoint deficient (*rad53-11*) background (Tourriere, Versini et al. 2005). This suggests that while the FPC is necessary to maintain replisome integrity at programmed RFBs (as discussed in 1.2.3.1), the FPC and Mrc1 are not required to stabilize replication forks that pause due to exogenous replication stress.

To prevent the generation of long stretches of ssDNA after the activation of the intra-S-phase checkpoint, the activity of the helicase must be limited. The FPC-Mrc1 complex has been shown to couple the activity of the CMG helicase and the polymerases when replication forks stall in the presence of HU. A ChIP-on-chip analysis of *S. cerevisiae* replisome components, showed that co-localisation of Cdc45 (i.e., CMG) with nascent DNA was uncoupled upon HU treatment in *tof1*Δ and *mrc1*Δ cells (Katou, Kanoh et al. 2003). A similar, though less severe, uncoupling was seen in *csm3*Δ (Bando, Katou et al. 2009). Taken together, this suggests that after polymerase stalling and checkpoint activation, the FPC-Mrc1 complex acts to prevent excessive unwinding of the duplex by coupling the polymerase and helicase activity.

Despite working as a complex to limit ssDNA accumulation after checkpoint activation, the recovery of replication after HU treatment in *tof1*Δ cells is much more efficient than that in *mrc1*Δ cells (Tourriere, Versini et al. 2005). Unlike *tof1*Δ, however, *mrc1*Δ cells do not inhibit late firing origins in HU (Tourriere, Versini et al. 2005, Hayano, Kanoh et al. 2011). This allows *tof1*Δ cells to fire the delayed

origins and complete replication after the HU block has been removed, thereby compensating for the replication fork defect (Tourriere, Versini et al. 2005). This highlights the importance of the intra-S-phase checkpoint dependent inhibition of late firing origins. It is important to note, however, that the precocious firing of origins in *mrc1*Δ has been recently suggested to be, at least in part, checkpoint independent (Matsumoto, Kanoh et al. 2017). This will be discussed in more detail in section 6.1.2.

1.2.3.2.3 Homologous recombination dependent replication fork restart

Collapsed replication forks can processed to allow restart using a homologous recombination (HR) mediated, but double strand break (DSB) independent, mechanism (Mizuno, Lambert et al. 2009, Lambert, Mizuno et al. 2010). This re-establishes a replication fork that can replicate DNA in a semi-conservative manner (Miyabe, Mizuno et al. 2015). Unlike canonical replication forks, HR-restarted forks are more error prone and replicate both leading and lagging strands using polymerase δ (Miyabe, Mizuno et al. 2015). Additionally, when HR-restarts replication it can lead to an increase in chromosomal rearrangements caused by non-allelic homologous recombination (Lambert, Mizuno et al. 2010).

Regions that do not finish DNA replication before mitosis form breaks on metaphase chromosomes. Loci that are statistically more likely to exhibit a break when the cells are subjected to mild replication stress are called common fragile sites (CFS) (Durkin and Glover 2007). In human cells, CFSs tend to associate with late replicating, AT-rich, origin-poor regions and large (>500 kb) transcriptional units (Glover, Wilson et al. 2017). It is possible that under increased replication stress, replication forks at CFSs stall and collapse more often and restart replication using error-prone HR-mediated fork restart, which can lead to chromosomal rearrangements.

It is important to note that fork stalling is not always beneficial or possible. When the replication fork encounters an RFB that blocks the unwinding activity of the CMG, e.g., intra-strand crosslinks (ICLs), replication forks collapse and the lesion is excised using a combination of nucleotide excision repair and HR (Zhang and Walter 2014). Disruption of the intra-S-phase checkpoint in *S. pombe* increased

resistance to ICLs, suggesting that fork stalling is not beneficial at ICLs and can even be detrimental (Lambert, Mason et al. 2003).

1.3 Heterochromatin formation and its impact on DNA replication

1.3.1 Roles of heterochromatin

DNA inside eukaryotic cells is wrapped around histone and non-histone proteins, forming a complex called chromatin. It was noted in as early as 1928 that certain chromosomal regions retain stains throughout the cell cycle, while others lose it when the chromosomes unpack in interphase. The differential staining resulted in chromosome-specific stripped patterns of "light" and "dark" bands, referred to as euchromatin and heterochromatin, respectively (the early cytological work on heterochromatin is reviewed in (Passarge 1979)). At a molecular level, heterochromatin is associated with various histone modifications (e.g., triple methylation of lysine 9 on histone H3, i.e., H3K9me3) and, in higher eukaryotes, DNA methylation (Volpe and Martienssen 2011). H3K9 methylation correlates strongly with repetitive DNA, low gene density and low transcriptional activity, suggesting a link between these chromosomal features and heterochromatin (Wang, Jia et al. 2016).

Heterochromatin in all eukaryotes can be subdivided into constitutive and facultative. Constitutive heterochromatin usually forms at the repetitive regions flanking centromeres (pericentromeric regions) and telomeres. Facultative heterochromatin, on the other hand, is cell type specific and has been linked to controlling developmentally regulated gene expression (Wang, Jia et al. 2016).

While it is most often associated with transcriptional silencing, heterochromatin has also been implicated in chromosome segregation and genome stability. In *S. pombe*, the absence of heterochromatin leads to a loss of sister chromatin cohesion and affects mitotic chromosome segregation (Hall, Noma et al. 2003). Heterochromatin formation at centromeres drives the incorporation of Cnp1 (*h*CENP-A), the centromere specific variant of histone H3 (Folco, Pidoux et al. 2008). Cnp1 has been shown to be important for the assembly of the kinetochore

in metazoans and yeast (Van Hooser, Ouspenski et al. 2001, Collins, Castillo et al. 2005, Regnier, Vagnarelli et al. 2005).

Heterochromatin has also been shown to inhibit HR at repetitive regions such as the *mat* locus in *S. pombe* (Grewal and Klar 1997) and rDNA repeats in *S. cerevisiae* (Guarente 2000). The maintenance of repetitive DNA cold to recombination is key to genome stability. It has been well characterised in both *S. cerevisiae* and *S. pombe* that nearby inverted repeats recombine spontaneously and frequently, leading to the formation of dicentric and acentric isochromosomes. Such forms of gross chromosomal rearrangements are unstable, do not segregate properly and lead to genome instability (Mizuno, Lambert et al. 2009, Paek, Kaochar et al. 2009), a hallmark of cancer (Negrini, Gorgoulis et al. 2010).

1.3.2 Formation and spreading of heterochromatin

In *S. pombe*, the formation of heterochromatin is regulated by an RNA interference (RNAi) pathway. RNAi was first described in *Caenorhabditis elegans* (*C. elegans*) as a process for targeted post-translational silencing of gene expression (Fire, Xu et al. 1998). The formation of heterochromatin, relies on the same proteins to target the activity of chromatin modification complexes (Volpe, Kidner et al. 2002). This section will discuss data derived mainly from experiments done in *S. pombe*. While RNAi and heterochromatin are both conserved, the link between them in higher eukaryotes is still elusive (Martienssen and Moazed 2015, Saksouk, Simboeck et al. 2015).

The process of heterochromatin formation (Fig. 1.3 A) is initiated by the transcription of short repeats by RNA Polymerase II (RNAP II) (Martienssen and Moazed 2015). In *S. pombe*, those repeats are found at the centromeres, telomeres, rDNA and *mat* locus (Cam, Sugiyama et al. 2005).The dsRNA fragments generated by RNAP II are cleaved into ~21 nucleotide long small interfering RNAs (siRNAs) by the catalytic activity of Dicer (Reinhart and Bartel 2002). The double stranded siRNAs are bound by argonaute protein Ago1, which in turn associates with Arb1 and Arb2 to form the ARC complex (Buker, Iida et al. 2007). Arb1 in the ARC complex inhibits the cleavage activity of Ago1, preventing the conversion of double stranded siRNAs to single stranded siRNAs (Buker, Iida et al. 2007). Ago1 bound to the duplex siRNAs eventually dissociates from ARC and binds to Tas3 and Chp1,

forming the RITS complex (Verdel, Jia et al. 2004, Irvine, Zaratiegui et al. 2006, Buker, Iida et al. 2007). The cleavage activity of Ago1 is not inhibited in RITS, allowing the formation of single stranded siRNAs. The siRNAs direct the RITS complex to loci from which they were transcribed by Watson-Crick base pairing (Irvine, Zaratiegui et al. 2006, Buker, Iida et al. 2007).The reason for the initial inhibition of Ago1 activity by ARC is unclear but it has been shown that unregulated Ago1 cleavage leads to the loss of RNAi mediated heterochromatin formation in *S. pombe* (Buker, Iida et al. 2007).

The RITS complex recruits the histone modification CLRC complex (Fig. 1.3 B) (Irvine, Zaratiegui et al. 2006), which consists of the H3K9 methyltransferase Clr4, H3K4 demethylase Lid2, ubiquitin E3 ligase Culin4 as well as accessory proteins Rik1, Clr7 and Clr8 (Hong, Villen et al. 2005, Horn, Bastie et al. 2005, Jia, Kobayashi et al. 2005, Li, Goto et al. 2005, Thon, Hansen et al. 2005). The methyltransferase activity of Clr4 methylates histone H3 (Ivanova, Bonaduce et al. 1998, Nakayama, Rice et al. 2001, Irvine, Zaratiegui et al. 2006). The H3K9me histone modification acts as a binding platform for Swi6 (Nakayama, Rice et al. 2001, Sadaie, Iida et al. 2004, Verdel, Jia et al. 2004, Cam, Sugiyama et al. 2005) the *S. pombe* the orthologue of heterochromatin protein 1 (HP1) and critical component of heterochromatin (Nakayama, Klar et al. 2000). The H3K9me/Swi6 platform recruits the SHREC complex, which reorganizes nucleosomes and has been proposed to mediate the formation of higher order chromatin structures (Sugiyama, Cam et al. 2007).

The RITS complex also acts to strengthen and spread the heterochromatin signal (Fig. 1.3 C). RITS recruits the RDRC (RNA-directed RNA polymerase complex), which amplifies the siRNAs, thereby propagating the heterochromatin signal (Halic and Moazed 2010). Heterochromatin spreading is driven by the chromodomain binding Chp1 component of the RITS complex, which binds directly to H3K9me marks (Partridge, Scott et al. 2002). The recruitment of CLRC to the new site of RITS binding spreads the heterochromatin signal around the initial H3K9 nucleation site (Verdel, Jia et al. 2004). In addition to RITS activity, the deacetylation of H4K16 by Sir2 was shown to increase nucleosome compaction

and proposed to facilitate the spreading of the H3K9 signal by Clr4 (Wang, Tadeo et al. 2013).

In *S. pombe*, boundaries around constitutive heterochromatin regions are established by Bdf2, which protects the H4K16 acetylation from Sir2 activity (Wang, Tadeo et al. 2013). Bdf2 is recruited to boundary sites by JmjC domain containing protein Epe1 (Zofall and Grewal 2006, Wang, Tadeo et al. 2013). The exact mechanism of Epe1 activity is still elusive. Epe1 has been widely described as an anti-silencing factor (Ayoub, Noma et al. 2003, Zofall and Grewal 2006, Braun, Garcia et al. 2011, Zofall, Yamanaka et al. 2012, Wang, Tadeo et al. 2013) and its role dependent on its JmjC domain (Ayoub, Noma et al. 2003), despite it not having any histone de-acetylase activity (Tsukada, Fang et al. 2006). The heterochromatin destabilizing activity of Epe1 also cannot be fully explained by its interaction with Bdf2. While heterochromatin spreading in *epe1*Δ*bdf2*Δ is similar to that of *bdf2*Δ, the effect of *epe1*Δ alone is consistently greater than that seen in *bdf2*Δ (Wang, Tadeo et al. 2013). This suggests that while the activity of Bdf2 in boundary formation is dependent on Epe1, Epe1 may have additional Bdf2-independent and JmjC-dependent boundary formation roles.

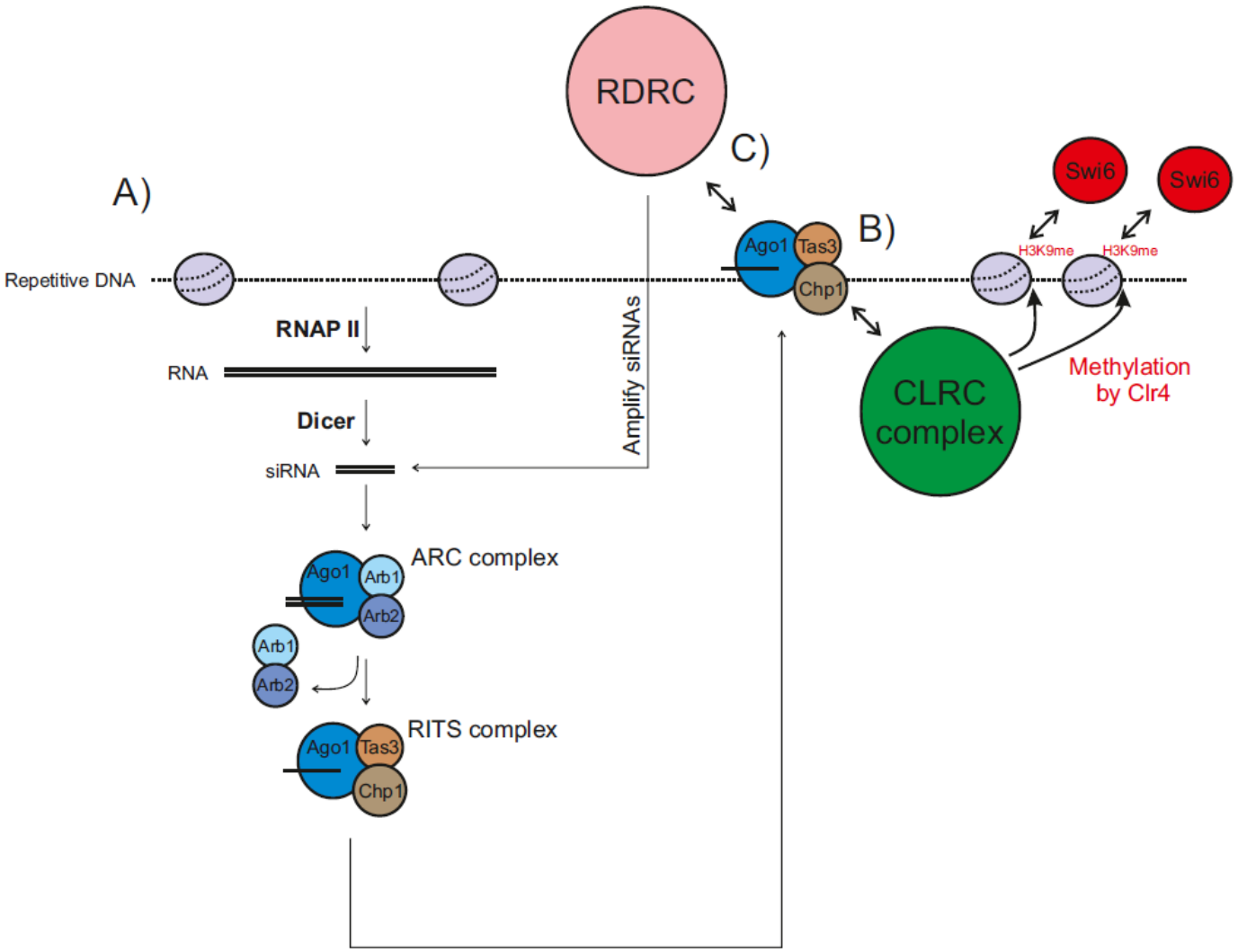

RDRC
C)
A)
B)
Swi6
Swi6
Ago1
Tas3
Chp1
H3K9me
H3K9me
Repetitive DNA
RNAP II
RNA
Dicer
siRNA
Amplify siRNAs
CLRC complex
Methylation by Clr4
ARC complex
Ago1
Arb1
Arb2
Arb1
Arb2
RITS complex
Ago1
Tas3
Chp1

Fig. 1.3 - **Formation and spreading of heterochromatin in *S. pombe***

A) RNA polymerase II (RNAP II) generates long RNA transcripts which are cleaved into ~21 bp long siRNA by Dicer. The double stranded siRNAs are shuttled to the RITS complex by ARC, where the Ago1 activity is inhibited. Ago1 slices the double stranded siRNAs when it is associated with the RITS complex. The single stranded siRNAs direct the RITS complex to the repetitive region from which they were transcribed.

B) The RITS complex recruits the CLRC complex, which contains the Clr4 methyltransferase. Clr4 methylates lysine 9 on histone H3 (nucleosome shown as a purple circle). The H3K9 methylation recruits Swi6, which is necessary for the formation of heterochromatin. RITS can also associate with H3K9 modification generated by CLRC, spreading the signal around the initial site of nucleation (not shown in the figure for simplicity)

C) The siRNA signal is amplified by RDRC (RNA-dependent RNA polymerase complex), which is recruited by RITS.

Adapted from (Martienssen and Moazed 2015).

1.3.3 Replicating heterochromatin

Chromatin is disassembled in front of replication forks and the histones are recycled into the replicated DNA to maintain the epigenetic marks. The details of this complex process may vary between organisms and are the subject of ongoing research.

Analysis of DNA replication foci in live mammalian cells showed that replicating chromatin is de-condensed at sites of replication (Kuipers, Stasevich et al. 2011). It has been proposed that the phosphorylation of linker histone H1 by CDK can also promote chromatin de-compaction in human cells (Alabert and Groth 2012). Additionally, a number of histone chaperones, some of which travel with the replisome, directly affect the nucleosomes ahead of the replication fork.

The H2A-H2B histone chaperone complex FACT, which travels with the replisome (Gambus, Jones et al. 2006), was shown to promote fork progression by directly disassembling nucleosomes (Abe, Sugimura et al. 2011). In human cells, histone H3-H4 chaperone Asf1 is linked to Mcm2-7 via a histone bridge and promotes DNA unwinding at sites of replication (Groth, Corpet et al. 2007). The current model of chromatin reassembly suggests that after depleting nucleosomes ahead of the fork, Asf1 and FACT associate with the newly formed naked DNA to deposit the nucleosomes and maintain epigenetic information (Rowlands, Dhavarasa et al. 2017).

In *S. pombe*, both FACT (Lejeune, Bortfeld et al. 2007) and Asf1 (Tanae, Horiuchi et al. 2012) were implicated in maintaining the integrity of heterochromatin at centromeres and overall genome stability. In addition to nucleosome remodellers at the replisome, other remodelling complexes are also conserved in *S. pombe*. One example is the chromatin remodeller Ino80, which has been shown to promote the exchange of histone H3 for histone H3 variant Cnp1 at the centromere (Choi, Cheon et al. 2017). The highly conserved presence of chromatin remodellers suggests that histone remodelling and recycling is conserved and acts upstream of replication forks in *S. pombe*. Despite this, heterochromatin in *S. pombe* has been linked to impeding replication fork progression. A study mapping the genome wide distribution of phosphorylated histone H2A (ϒH2A, a marker of DNA damage)

showed an enrichment of γH2A around sites of constitutive heterochromatin (Rozenzhak, Mejia-Ramirez et al. 2010), suggesting that it may be a form of an endogenous RFB, similarly to *RTS1* or rDNA barriers.

Heterochromatin has also been implicated in the DNA replication timing program (described in 1.4). Heterochromatin was first shown to replicate late in the 1960s (Lima-de-Faria and Jaworska 1968) and has since been corroborated in multiple organisms (Gilbert 2002, Zink 2006). An important exception to the late replication of heterochromatin is the early replication of the centromeres and *mat* locus in *S. pombe* (Kim, Dubey et al. 2003). The early replication is driven by the interaction between Swi6 and Dfp1, the regulatory subunit of DDK kinase (Hayashi, Takahashi et al. 2009), whose activity is needed for the activation of Mcm2-7 at the pre-RC and the binding of initiation factors to form the pre-IC (discussed in 1.2.1). It is unclear why this effect is not seen at telomeres, which are also encompassed in Swi6-dependent heterochromatin (Ekwall, Javerzat et al. 1995). It has also been shown that Swi6 interacts with Cdc18 (*sc*Cdc6) (Li, Chretien et al. 2011), suggesting that Swi6 may have more complex DDK independent roles in regulating origin firing.

1.4 DNA replication timing

The process of DNA replication in eukaryotic cells does not start simultaneously at every origin during S-phase. Instead, individual origins fire in a defined temporal order and are often described as "early" or "late" firing (Rhind and Gilbert 2013). These descriptions, misleadingly, suggest that each "early" or "late" origin will fire at the same time point in every cell in a population, i.e., that the process of origin firing is deterministic. This, however, is not the case. In both yeast and metazoans, origin firing has been shown to be stochastic (Lebofsky, Heilig et al. 2006, Patel, Arcangioli et al. 2006, Czajkowsky, Liu et al. 2008, Kaykov and Nurse 2015), i.e., in a population, the probability that an origin fires during S-phase varies between cells. Despite this stochasticity, large genomic segments (replication domains) replicate in the same temporal order in a population. The order with which replication domains are replicated in a cell is called the global replication timing (RT) program. It has been reported that the global RT program is deregulated in

cancer (Donley and Thayer 2013), although it has not been established whether it is a cause or consequence of it. The exact physiological role of the global RT program has yet to be defined.

The global RT program of a population can be measured using various direct and indirect methods and is presented as a global RT profile (Rhind and Gilbert 2013, Rivera-Mulia and Gilbert 2016).

1.4.1 Reconciling stochasticity of origin firing with a defined temporal order

The stochastic nature of origin firing can be seen as paradoxical with the reproducible global RT profiles. The two concepts were described in great detail in *S. cerevisiae* and were reconciled using mathematical models. This section will describe these studies, explore the mathematical models built on them and the impact of their assumptions.

A comprehensive single molecule DNA combing study showed that origin firing in *S. cerevisia*e is stochastic (Czajkowsky, Liu et al. 2008). The same study recapitulated previously published global RT profiles (Raghuraman, Winzeler et al. 2001), proving that the intrinsic stochastic nature of origin firing can lead to a reproducible RT profile.

Stochastic origin firing and a constant global RT program were reconciled in a simple *in silico* simulation (Rhind, Yang et al. 2010), which modulated origin firing efficiencies during S-phase. This simulation was built on two main assumptions. Firstly, origins had different relative probabilities of being fired during S-phase, therefore, origins with low probability of firing did not often fire early in S-phase. Secondly, the probability of firing of all unreplicated origins increased towards the end of S-phase (Yang, Rhind et al. 2010). This assumption was incorporated to overcome the so-called "random gap problem" (Rhind 2006), the concept that the stochasticity of origin firing can occasionally lead to large unreplicated gaps. Increasing the probability of origin firing towards the end of S-phase, allowed all origins that had not been already passively replicated to fire and complete DNA replication. Biologically, the increase in the relative probability of firing was explained by the increase in the concentration of unused limiting factors towards the end of replication (Rhind, Yang et al. 2010). As the number of origins that have

not fired decreases during S-phase, the relative concentration of the limiting factors available increases, allowing more efficient origin firing.

Later the same year, a more comprehensive model of DNA replication in *S. cerevisiae* allowed the reconstruction of the RT profiles from microarray data (Yang, Rhind et al. 2010). The model differentiated between "observed" and "potential" efficiencies of origins, where the latter was the predicted efficiency of an origin if it had not been passively replicated during S-phase. Both the observed and potential efficiencies, however, were extracted from firing-time distributions. This causally linked the probability of an origin firing during S-phase with the time at which it fired. Unsurprisingly, the model predicted that origins with the lowest "potential efficiency" fire late in S-phase. This correlated well with the earlier model (Rhind, Yang et al. 2010) whose critical assumption was that inefficient origins do not fire in early S-phase.

The (Yang, Rhind et al. 2010) model accurately reconstructed the temporal pattern across all chromosomes in *S. cerevisiae* (Yang, Rhind et al. 2010). Based on the genome wide reconstruction of RT data, it has been proposed that modulating origin firing across large domains is the cause of global RT programs. An important caveat of the model, however, is that the analysis was limited to the reconstruction of RT profiles over ~1.5 Mb (the length of the largest chromosome in *S. cerevisiae*). This is considerably smaller than human chromosomes, where the smallest chromosome (Chromosome 21) is over 46 Mb long.

The connection between firing an origin efficiently and the surrounding loci being replicated early in S-phase is clear. In organisms with homogenous RT profiles, i.e., one without clear early and late replicating domains on the same short chromosome, changes in origin firing efficiency could affect the global RT profile. The models described above do well in predicting such patterns It is unclear, however, whether this can be extrapolated to explain genome wide RT patterns in higher eukaryotes, whose large chromosomes often have replication timing domains the size of *S. cerevisiae* chromosomes.

1.4.2 Replication timing domains in the context of genome organisation

The advent of next generation sequencing (NGS) has allowed high resolution and comprehensive analysis of RT profiles. Independently, novel NGS techniques have elucidated the 3D nuclear organisation of metazoan and yeast genomes. This section will describe the experiments that led to the discovery of replication timing and structural chromosomal domains as well as their relationship in metazoans.

Early RT profiles of mouse embryonic stem cells (mESCs) revealed that the genome is replicated in Mb long replication domains, separated by transition zones (Hiratani, Ryba et al. 2008). As described in 1.4.1, the average RT across the replication domains is constant on a population level, despite the stochastic nature of origin firing. The replication domains and transition zones linking them are, therefore, often referred to as CTRs (constant timing regions) and TTRs (transition timing regions), respectively (Rhind and Gilbert 2013, Pope, Ryba et al. 2014). A comparison of mESCs that represented 10 stages of early mouse development showed that ~45% of the genome changes its relative replication time in S-phase during cell differentiation. These changes were also shown to affect transcription (Hiratani, Ryba et al. 2010). The RT profiles across different cells lines showed that the regions encompassed in CTRs and TTRs were mostly conserved and only the relative time at which the CTRs were replicated was affected (Hiratani, Ryba et al. 2010). A comparison of RT profiles from human and mouse ESCs showed considerable conservation across regions of synteny, suggesting an evolutionarily conserved role for replication domains (Ryba, Hiratani et al. 2010). Additionally, human and mouse chromosomes with rearrangements and translocations showed a divergence from a wild type RT profile only at the position of the rearrangements (Yaffe, Farkash-Amar et al. 2010, Pope, Chandra et al. 2012). Taken together, these data suggest that replication timing domains are conserved and self-associating units and that the relative timing of their replication may be linked to gene expression.

Large regions of human (Guelen, Pagie et al. 2008), *C. elegans* (Ikegami, Egelhofer et al. 2010) and *D. melanogaster* (Pickersgill, Kalverda et al. 2006) genomes were shown to associate with the nuclear lamina (the underlying mesh of the nuclear envelope) using DamID methods (van Steensel, Delrow et al. 2001). Briefly, a DNA

methyltransferase is fused to a component of the nuclear lamina (Gruenbaum and Foisner 2015), methylating any loci that contact it. These regions are subsequently isolated, sequenced and mapped to reference genomes (van Steensel, Delrow et al. 2001). This allows the generation of genome wide maps that describe the lamina associated domains (LADs) in a cell. Loci found within LADs correlated with a repressive heterochromatin environment and low gene expression (Pickersgill, Kalverda et al. 2006, Guelen, Pagie et al. 2008, Ikegami, Egelhofer et al. 2010).

In recent years, data from genome wide chromosome conformation capture (Hi-C) techniques further enhanced our understanding of the roles that the compartmentalisation of DNA in the nucleus plays in DNA replication. Hi-C data produce contact frequency maps, from which the 3D distribution of the genome can be extrapolated (Lieberman-Aiden, van Berkum et al. 2009). Briefly, interacting loci are crosslinked, the genome is digested into small fragments and the loci are sequenced. An early low resolution study of large scale interactions in the human genome revealed a spatial segregation of open and closed chromatin (into "A" and "B" domains, respectively) (Lieberman-Aiden, van Berkum et al. 2009). A related Hi-C method described the spatial organisation of 4.5 Mb region of the mouse Chromosome X at a very high resolution (Nora, Lajoie et al. 2012). Both active and inactive X Chromosomes were organised into 0.2 - 1 Mb long topologically associated domains (TADs), whose boundaries correlated well with those of LADs. Expression of genes whose promoters were found within the same TAD correlated very strongly with each other (Nora, Lajoie et al. 2012). This suggested that TADs, similarly to replication domains, were self-associating regions linked to gene expression. High resolution Hi-C of mESCs and hESCs revealed genome wide TAD formation, showing that they are key features of genome organisation (Dixon, Selvaraj et al. 2012). The organisation of TADs was also shown to be highly conserved over regions of synteny between mouse and human genomes (Dixon, Selvaraj et al. 2012), which is highly reminiscent of replication timing domains (Hiratani, Ryba et al. 2010). The boundaries of a number of the TADs identified were shown to correlate well with boundaries of LADs, "A" and "B" chromatin and early and late replicating domains (Dixon, Selvaraj et al. 2012). This comprehensive analysis showed for the first time that all

of these independently identified domains are linked. More recently, TAD and replication domain boundaries in 31 different mouse and human cell types were mapped with a one-to-one correlation (Pope, Ryba et al. 2014), proving a conclusive link between the two. To date, however, no causal relationship between these elements has been described. A summary of the relationship between chromosomal domains, replication timing domains and LADs is shown in Fig. 1.4.

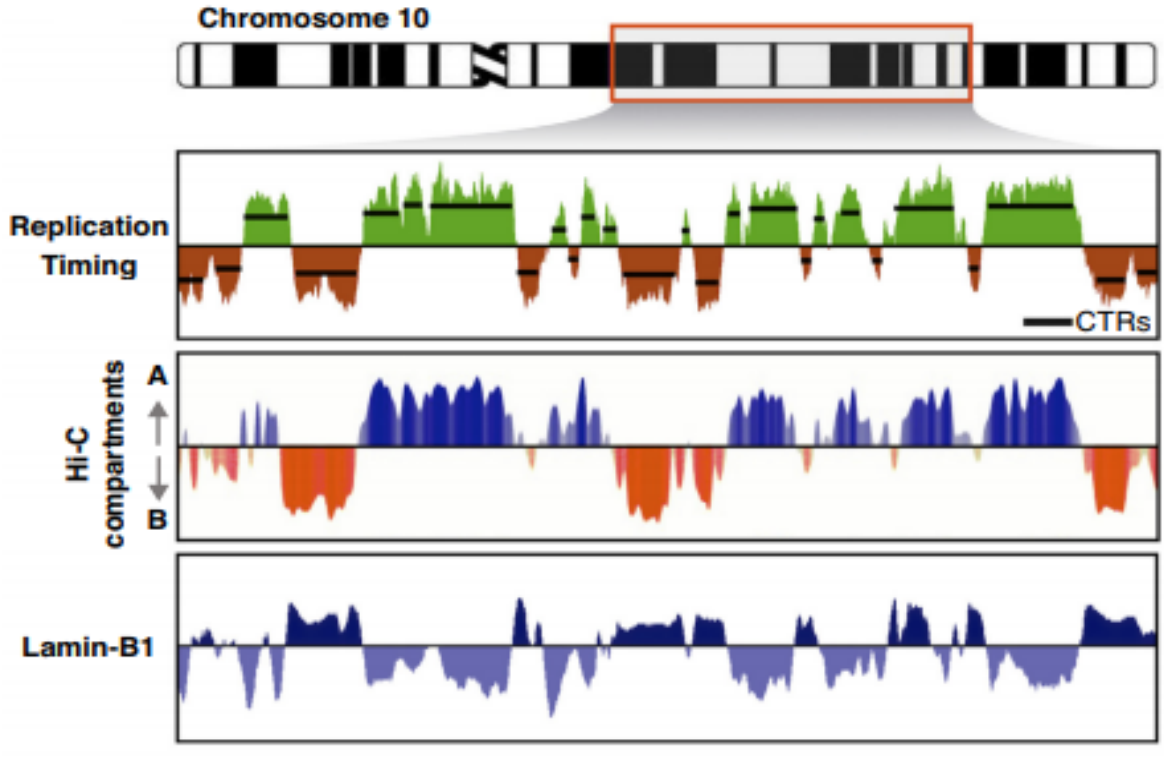


Fig 1.4 - **Relationship between metazoan chromosomal domains, replication timing domains and LADs**

A replication timing profile of a 50 Mb region on human Chromosome 10 from IMR90 fibroblasts. The segments of chromosomes with uniform replication timing (constant timing regions, CTRs) were aligned with Hi-C interaction compartments and LaminB1 contact maps.

Adapted from (Rivera-Mulia and Gilbert 2016).

1.4.3 Establishing the replication timing program in metazoans

The indication that metazoans establish their replication timing program in G_1 was first reported nearly 20 years ago (Dimitrova and Gilbert 1999). In this study, Chinese hamster nuclei were isolated at different stages of G_1 and introduced into a cell free system (*Xenopus* egg extracts) to complete replication. G_1 nuclei harvested after the loading of Mcm2-7 onto origins and at least 1 hour after nuclear envelope formation maintained a wild type RT program. This cell cycle point was called the replication timing decision point (TDP) (Dimitrova and Gilbert 1999). It was also reported that the repositioning of chromosomal domains in G_1 was associated with the establishment of the RT program (Dimitrova and Gilbert 1999), linking for the first time chromosomal structures and replication timing. More recently, a Hi-C analysis of a mouse cell line showed that TADs are established in the same time frame as the TDP (Dileep, Ay et al. 2015), further demonstrating the connection between TAD formation and the RT program.

It is still unclear what drives the establishment of TADs. TAD formation at the mouse Chromosome X inactivation centre was not affected in cell lines lacking H3K9 methylation (Nora, Lajoie et al. 2012). Additionally, TADs shared between pluripotent and terminally differentiated cells were shown to act as boundaries to H3K9 methylation (Dixon, Selvaraj et al. 2012), despite the two types of cells having different H3K9me distributions. These data suggest that TAD formation is not driven by heterochromatin, rather TADs may act to demarcate potential heterochromatin boundaries. TAD boundary regions are enriched for insulator binding protein CTCF and cohesin, however, the binding of CTCF and/or cohesin alone is not sufficient for boundary formation (Dixon, Selvaraj et al. 2012, Nora, Lajoie et al. 2012). In mESCs, Rif1 was shown to associate with large chromosomal regions forming RADs (Rif1 associated domains) (Foti, Gnan et al. 2016). RADs were shown to be late replicating domains, correlate with LADs and restrict interactions between replication timing domains in G_1 (Foti, Gnan et al. 2016). This suggested a role for Rif1 in establishing the replication timing profile by directly affecting nuclear architecture.

1.4.4 Genome organisation and global replication timing program in yeast

Similarly to higher eukaryotes, yeast genomes are organised into replication timing domains (Raghuraman, Winzeler et al. 2001, Czajkowsky, Liu et al. 2008, Daigaku, Keszthelyi et al. 2015). As discussed in 1.4.1, in *S. cerevisiae*, replication timing domains can span whole chromosomes, due to their relatively small size (Yang, Rhind et al. 2010). The *S. pombe* replication timing domains span ~ 2 Mb and timing transition zones are best visible on Chromosomes 1 and 2 (Daigaku, Keszthelyi et al. 2015). In comparison to metazoans, however, the correlation between self-associating chromosomal domains and replication timing domains have not been well described in yeast.

Initial Hi-C reports suggested that while TADs are highly conserved across eukaryotes, they are not present in *S. cerevisiae* (Duan, Andronescu et al. 2010, Sexton and Cavalli 2015). A modified Hi-C approach identified ~5 kb TAD-like chromosomal interaction domains (CIDs) which were linked to transcription (Hsieh, Weiner et al. 2015). More recent Hi-C analysis identified larger ~200 kb long TAD-like domains (Eser, Chandler-Brown et al. 2017), which were enriched for origins with similar relative firing times (Eser, Chandler-Brown et al. 2017).

50 - 100 kb long TAD-like structures (called "globules") have been identified in *S. pombe* (Mizuguchi, Fudenberg et al. 2014). Similarly to TADs of higher eukaryotes (Dixon, Selvaraj et al. 2012, Nora, Lajoie et al. 2012), *S. pombe* globule formation is dependent on cohesin and not on heterochromatin (Mizuguchi, Fudenberg et al. 2014). No correlation with replication domains was reported but loss of cohesin did affect transcription (Mizuguchi, Fudenberg et al. 2014). A polymer model incorporating Hi-C and microscopy data from *S. pombe* revealed a nuclear compartmentalisation similar to that seen in higher eukaryotes (Pichugina, Sugawara et al. 2016). The model showed that early firing origins cluster in the centre of nucleus, while late firing origins are more likely to be associated with the nuclear periphery. Interestingly, loci that had euchromatic and heterochromatic marks in wild type cells, did not change their nuclear positions in *clr4*Δ cells (Pichugina, Sugawara et al. 2016). This suggests that, similarly to metazoans (Dixon, Selvaraj et al. 2012, Nora, Lajoie et al. 2012), where heterochromatin does

not drive TAD formation, heterochromatin does not influence the global distribution of genomic DNA in *S. pombe*.

The global architecture of yeast genomes may be fundamentally different to that of mammalian ones. For example, a polymer model predicted that only 30-40% of the *S. cerevisiae* genome is covered in topological loops (Schalbetter, Goloborodko et al. 2017), compared to 100% predicted for mammalian chromosomes (Naumova, Imakaev et al. 2013). The RT program may, therefore, have different drivers. While also implicated in the global RT program in yeast (Hayano, Kanoh et al. 2012, Hiraga, Alvino et al. 2014, Kanoh, Matsumoto et al. 2015), in *S. pombe*, Rif1 is not very abundant (Marguerat, Schmidt et al. 2012) and can only be visualised when over-expressed (Zaaijer, Shaikh et al. 2016). It is, therefore, unlikely that Rif1 plays a role in the structural organisation of the genome, as has been reported for metazoans (Foti, Gnan et al. 2015).

Other factors have been implicated in the regulation of origin firing and replication timing in yeast. In *S. cerevisiae*, the conserved forkhead transcription factors (TF) Fkh1 and Fkh2 have been shown to bind early replicating origins and drive their clustering in replication factories (Knott, Peace et al. 2012), independently of their roles as TFs (Ostrow, Kalhor et al. 2017). The TF activity of *S. pombe* Fkh2 was implicated in both mitosis (Bulmer, Pic-Taylor et al. 2004) and meiosis (Alves-Rodrigues, Ferreira et al. 2016). There are, however, no reports on the role of Fkh2 as a regulator of the global RT program. On the other hand, Mrc1, the regulator of the intra-S-phase checkpoint described in 1.2.3.2, was reported to bind early replication origins and affect the RT profile of *S. pombe*, in a checkpoint independent manner (Hayano, Kanoh et al. 2011, Matsumoto, Kanoh et al. 2017) (discussed in more detail in 6.1.2). Taken together, these data suggest that while some of the proteins and processes are conserved between yeast and metazoans, the pathways determining the global RT program in yeast may not be.

1.5 Summary and aims

DNA replication is a highly regulated process whose fidelity is crucial to genome stability. *S. pombe* is a well-studied model organism whose DNA replication dynamics have been extensively described. Numerous aspects of the DNA replication dynamics in *S. pombe*, however, remain elusive. The work in thesis addresses two aspects of DNA replication using novel NGS methods. Firstly, we investigated the impact of heterochromatin on replication fork progression. We analysed replication forks and origins firing in constitutive and facultative heterochromatin in a number of genetic backgrounds. Secondly, we examined the mechanism controlling the global RT program in *S. pombe* by analysing the impact of global origin usage, heterochromatin and the nuclear distribution of chromatin.

Chapter 2

2.1 Materials

2.1.1 Growth media and agar plates

- **Yeast Extract (YE) media**

  In 1 L - 5 g yeast extract, 30 g glucose, 0.8 g adenine, 0.4 g each of leucine, uracil, histidine and arginine. Autoclaved

- **YE agar (YEA) plates**

  6.25 g Bacto Agar was autoclaved in 500 mL of YE media and poured into sterile petri dishes.

- **YE with peptone (YEP)**

  YE media supplemented with 20 g/L peptone.

- **4X Edinburgh minimal media (EMM)**

  In 1 L - 50 mL 20X EMM salts, 25 mL 20% NH_4Cl, 25 mL 0.4 M $NaHPO_4$, 50 mL 40% glucose, 1 mL 1000X vitamins, 100 µl 10,000X trace elements. Filter sterilised using a Steritop™ 0.22 µm 1 L filter unit (Merck Millipore).

  20X EMM salts

  In 1 L - 61.2 g $C_8H_5KO_4$, 20 g KCl, 21.4 g $MgCl_2.6H_2O$, 0.2 g Na_2SO_4, 0.26 g $CaCl_2.2H_20$

  1000X vitamins

  In 100 mL - 100 mg pantothenic acid, 1 g nicotinic acid, 1 g inositol, 1 mg biotin

  10,000X trace elements

  In 100 mL - 500 mg H_3BO_3, 400 mg $MnSO_4$, 400 mg $ZnSO_4.7H_2O$, 200 mg $FeCl_2.6H_2O$, 150 mg Na_2MoO_4, 100 mg KI, 40 mg $CuSO_4.5H_2O$, 1 g citric acid

- **EMM agar plates**

  10 g Formedium Bacto Agar was autoclaved in 300 mL ultra-pure water. 100 mL of 4X EMM was added to molten agar and supplemented with adenine, leucine and/or uracil, as needed (final concentration of supplements 1 g/L).

- **Lysogeny broth (LB) media**

  In 1 L - 10 g tryptone, 5 g yeast extract, 5 g NaCl. Autoclaved

- **LB agar plates**

  6 g Bacto Agar was autoclaved in 500 mL of LB media and poured into sterile petri dishes.

- **Extremely low nitrogen (ELN) agar plates**

  In 1 L - 27.3 g Formedium EMM Broth (without nitrogen), 0.05 g ammonium chloride, 0.2 g adenine, 0.1 g each of leucine, uracil, histidine and arginine, 25 g Bacto Agar. Autoclaved and poured into sterile petri dishes.

2.1.2 Drugs used for genetic selection

Drug	Final concentration
Nourseothricin sulphate (NAT)	100 µg/ml
Geneticin dislulphate (G418)	200 µg/ml
5-fluoroorotic acid (5-FOA)	1 mg/ml
Blasticidin (BDS)	15 µg/ml
Ampicillin sodium salt (AMP)	100 µg/ml

2.1.3 Buffers

- NIB

In 80 mL - 1.05 g MOPS, 1.47 g KAc, 17 mL 100% glycerol, 200 µl $MgCl_2$. Adjust to pH 7.2 with 10 M KOH. Filter sterilised using a Steritop™ 0.22 µm 1 L filter unit (Merck Millipore).

- 5X TBE buffer

In 1 L - 54 g Tris base, 27.2 g boric acid, 20 mL 0.5M EDTA.

- 10X TE

In 1 L - 19.7 g Tris HCl, 4.7 g EDTA. Adjust to pH 8.0 with NaOH and autoclave.

- 50 mM Citrate phosphate buffer

In 1 L - 7.10 g Na_2HPO_4 (anhydrous), 11.50 g citric acid. Adjust pH to 5.6 and autoclave.

- CSE

In 1 L - 218.6 g sorbitol, 80 µl 0.5M EDTA in 50 mM citrate phosphate buffer. Filter sterilised using a Steritop™ 0.22 µm 1 L filter unit (Merck Millipore).

2.1.4 Strain list

Table 2.1- **List of strains used in this study**

Name	Mating type	Genotype	Description	Experiments used in
AMC501	*h-*	*ade6-704 ura4-D18 leu1-32*	Wild-type	Pu-Seq
IM655	*h-*	*ade6-704 ura4-D18 leu1-32 rnh201::KanMX cdc20-M630F*	"wild type" Pol ε Pu-Seq strain - Pol ε mutation in *rnh201*Δ background	
IM856	*h-*	*ade6-704 ura4-D18 leu1-32 rnh201::KanMX cdc6-L591G*	"wild type" Pol δ Pu-Seq strain - Pol δ mutation in *rnh201*Δ background	
YKP017	*h-*	*ade6-704 ura4-D18 leu1-32 rif1::BSD rnh201::KAN cdc20-M630F*	*rif1*Δ Pol ε Pu-Seq strain	
YKP019	*h-*	*ade6-704 ura4-D18 leu1-32 rif1::BSD rnh201::KAN cdc-6L591G*	*rif1*Δ Pol δ Pu-Seq strain	
PLK053	*h+*	*ura4-D18 leu1-32 rnh201::KanMX taz1::NAT cdc20-M630F*	*taz1*Δ Pol ε Pu-Seq strain	
PLK054	*h-*	*ura4-D18 leu1-32 rnh201::KanMX taz1::NAT cdc6-L591G*	*taz1*Δ Pol δ Pu-Seq strain	
PLK073	*h-*	*ura4-D18 leu1-32 man1::NAT rnh201::KANMX cdc20-M630F*	*man1*Δ Pol ε Pu-Seq strain	
PLK074	*h-*	*ura4-D18 leu1-32 man1::NAT rnh201::KANMX cdc6-L591G*	*man1*Δ Pol δ Pu-Seq strain	
PLK083	*h+*	*ura4-D18 leu1-32 epe1::NAT rnh201::KANMX cdc6-L591G*	*epe1*Δ Pol ε Pu-Seq strain	
PLK084	*h+*	*ura4-D18 leu1-32 epe1::NAT rnh201::KANMX cdc20-M630F*	*epe1*Δ Pol δ Pu-Seq strain	
YKP038	*h-*	*ade6-704 ura4-D18 leu1-32 rnh201::KanMX rif1-PP1 cdc20-M630F*	Rif1-PP1 Pol ε Pu-Seq strain	
YKP039	*h-*	*ade6-704 ura4-D18 leu1-32 rnh201::KanMX rif1-PP1 cdc6-L591G*	Rif1-PP1 Pol δ Pu-Seq strain	
YKP048	*h-*	*ade6-704 ura4-D18 leu1-32 rnh201::KanMX rif1-7A cdc20-M630F*	Rif1-7A Pol ε Pu-Seq strain	
YKP049	*h-*	*ade6-704 ura4-D18 leu1-32 rnh201::KanMX rif1-7A cdc6-L591G*	Rif1-7A Pol δ Pu-Seq strain	
YKP026	*smt0 h-*	*ade6-704 ura4-D18 leu1-32 smt0 rnh201::KanMX swi6::NAT cdc20-M630F*	*swi6*Δ Pol ε Pu-Seq strain	
YKP024	*smt0 h-*	*ade6-704 ura4-D18 leu1-32 smt0 rnh201::KanMX swi6::NAT cdc6-L591G*	*swi6*Δ Pol δ Pu-Seq strain	
YKP036	*h-*	*ade6-704 ura4-D18 leu1-32 rnh201::KanMX clr4::NAT cdc20-M630F*	*clr4*Δ Pol ε Pu-Seq strain	

YKP037	*h-*	*ade6-704 ura4-D18 leu1-32 rnh201::KanMX clr4::NAT cdc6-L591G*	*clr4*Δ Pol δ Pu-Seq strain	
BAF59	*h+*	*ade6-704 ura4-D18 leu1-32 rif1::BSD*	*rif1*Δ	Elutri-Seq
BAF394	*h+*	*ade6-704 ura4-D18 leu1-32 his3-D1 rif1-PP1*	Rif1-PP1	
MS253	*h-*	*ura4-D18 leu1-32 mrc1::KanMX*	*mrc1*Δ	
PLK076	*h-*	*ade6-704 ura4-D18 leu1-32 loxP::rif1-mEOS::loxM3 Gar2-GFP*	Rif1 tagged with mEOS Gar2 tagged with GFP	Single molecule microscopy
PLK081	*h-*	*ade6-704 ura4-D18 leu1-32 loxP::rif1PP1-mEOS::loxM3 Gar2-GFP*	Rif1-PP1 tagged with mEOS Gar2 tagged with GFP	

2.2 Methods

2.2.1 General molecular cloning techniques

2.2.1.1 Polymerase chain reaction (PCR) methods

Colony PCR

A single colony was mixed into 25 µl double distilled H_20 (ddH_2O) and boiled at 95°C. The following master mix was prepared separately:

Component	**Volume for 25 µl reaction**
10X Taq buffer	5 µl
dNTPs (2 mM each)	5 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
ddH_2O	12.75 µl
Super-Therm Taq polymerase (Labmaster)	0.25 µl

25 µl of the master mix was added to the cells after 10 minutes at 95°C, for a total volume of 50 µl. The thermocycling conditions were:

95°C	2 minutes	
95°C	20 seconds	X29
Primer melting temperature -4°C	10 seconds	
72°C	30 seconds / kb	
72°C	10 minutes	
20°C	Hold	

PCR from genomic DNA

To amplify fragments for cloning, KOD hot start polymerase kit (Novagen) was used:

Component	**Volume for 50 µl reaction**
10X KOD buffer	5 µl
dNTPs (2 mM each)	5 µl
MgSO4 (25 mM)	4 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
DNA*	
ddH_2O	Make reaction up to 50 µl
KOD polymerase	1 µl

*100ng and 10ng of DNA was used to amplify fragments from genomic and plasmid DNA, respectively.

The thermocycling conditions were:

95°C	2 minutes	
95°C	20 seconds	X29
Primer melting temperature -4°C	10 seconds	
68°C	30 seconds / kb	
68°C	10 minutes	
20°C	Hold	

2.2.1.2 Restriction digests

Restriction enzymes were purchased from New England Biolabs (NEB). Restriction digests were performed according to the manufacturer's conditions.

2.2.1.3 Cloning

Cloning was carried out using NEBuilder® HiFi DNA Assembly kits (NEB). The vector was linearized using restriction enzymes and ran on 1% TBE agarose gel. The linearized fragment was excised using a scalpel and the DNA was extracted from the agarose gel using Macherey-Nagel Gel and PCR Clean-up Kit, following manufacturer's instructions. Next, the fragments that were to be cloned into the vector were amplified with overhangs, to comply with NEBuilder® HiFi DNA Assembly kits recommendations.

To clone 1 insert into a vector, 0.02 pmoles of vector were mixed with 0.04 pmoles of insert. 5 µl of NEBuilder HiFi DNA Assembly Master Mix was added and the total volume was taken up to 10 µl using ultra-pure water. The reaction was incubated in a thermocycler for 15 minutes at 50°C.

To clone 3 inserts (where each insert was at least 200 bp long) into a vector, 0.05 pmoles of each insert was mixed and 5 µl of NEBuilder HiFi DNA Assembly Master Mix was added. The total volume was taken up to 10 µl (minus the volume of 0.01 pmoles of vector) using ultra-pure water. The reaction was incubated in a thermocycler for 15 minutes at 50°C and then for a further 15 minutes on ice. 0.01 pmoles of vector was added and the reaction was incubated in a thermocycler for 1 hour at 50°C.

2.2.2 General *E. coli* cell biology techniques

2.2.2.1 *E. coli* transformation

40 µl of competent DH5α *E. coli* cells were thawed on ice. Cells were incubated with 2 µl of NEB HiFi Assembly product for 20 minutes on ice and heat shocked for 1 minute at 42°C. After heat shocking, the cells were incubated on ice for a further 5 minutes. 700 µl of LB broth was added, the cells were left to grow for 1 hour at 37°C and were then plated onto LB agar plates, supplemented with ampicillin for selection.

2.2.2.2 Extraction of plasmids from *E. coli*

Small and large scale plasmid extractions were done using Macherey-Nagel mini and midi prep kits, respectively, following manufacturer's instructions.

2.2.3 General *S. pombe* cell biology techniques

2.2.3.1 Crossing *S. pombe* strains and random spore analysis

S. pombe h+ and *h-* strains were mixed in 50 µl of ddH_2O on ELN agar plates. The patch was left to dry and the plate was incubated at 25°C for 3 days. A patch of crossed cells was resuspended in 1 mL of ddH_2O. 1 µl of β-galactosidase from *Helix pomentia* (Roche) was added and the cells were left to rotate at room temperature overnight. The following day, the spores were counted using a haemocytometer, 1000 spores were plated onto YEA and incubated at 30°C for 3 days. The resultant colonies were replica plated onto YEA supplemented with drugs or EMM to check for selection markers.

2.2.3.2 *S. pombe* transformation

Cultures were grown overnight to a final density of $1x10^7$ cells/mL (counted using haemocytometer). $1x10^8$ cells were washed with ddH_2O, pelleted and resuspended in 1 mL of LiAc-TE. 2 µl of boiled salmon sperm DNA (10 mg/mL) and 500 ng of plasmid DNA were added and the cells were incubated at room temperature for 10 minutes. 260 µl of 40% (w/v) PEG (dissolved in LiAc-TE) was added and the cells were incubated for 1 hour at 30°C. 43 µl of DMSO was added and the cells were mixed and heat shocked for 5 minutes at 42°C. Immediately after, the cells were pelleted, washed with 1 mL of ddH_2O, pelleted again and resuspended in 500 µl of ddH_2O. 2 x 250 µl was plated out onto 2 YEA plates, which were incubated at 30°C until colonies appeared. The colonies were restreaked onto fresh YEA plates and replica plated onto YEA supplemented with drugs or EMM to check for selection markers.

2.2.3.3 Recombination mediated cassette exchange (RMCE)

All strains were made using non-essential gene replacement, as described by (Watson, Garcia et al. 2008).

Base strain construction

The *ura4* marker gene flanked by two incompatible lox sites - *loxM3* and *loxP* (collectively referred to as the '*ura4* cassette') was amplified from plasmid pAW1 using primers with 80 bp long overhangs, which were homologous to the regions flanking the desired deletion. The linear fragment was used to transform *ura- leu- S. pombe* strains, as described in 2.2.3.2 (using 10 µl of the unpurified PCR reaction). Recombination between the regions of homology in the genome and on the linear fragment replaced the targeted gene with the *ura4* cassette. Transformed cells were selected by screening for uracil prototrophy.

Plasmid construction

The genes of interest were amplified and cloned into plasmid pAW8 as described in 2.2.1.3. Plasmid pAW8 contains the *leu1* marker gene and the gene coding for the Cre recombinase.

Strain transformation and selection

The pAW8 based plasmids were used to transform the base strain, as described in 2.2.3.2. Transformed cells were selected by screening for uracil and leucine prototrophy. A single colony was used to inoculate 10 mL of YEA, which was grown overnight to saturation at 30°C. This allowed the recombination (catalysed by Cre) of the *loxP::gene::loxM3* construct on the plasmid with the *loxP::ura4::loxM3* construct in the base strain. 10^4 cells were plated out on YEA+5FOA and incubated at 30°C for 3 days, to screen for uracil auxotrophy. The resultant colonies were streaked out onto YEA and replica plated to check for leucine auxotrophy (loss of pAW8).

2.2.3.4 Genomic DNA extraction (for PCR)

Strains were grown in 10 mL of YE overnight at 30°C. 1 mL of the saturated culture was pelleted, resuspended in 1 mg/mL lyticase (Merrick and Fisher) in CSE and incubated at 37°C for 30 minutes. The cells were pelleted again and resuspended in

450 µl of 5X TE. 50 µl of 10% SDS was added and the sample was incubated at room temperature for 5 minutes. 150 µl of 5 M KAc was added and the sample was incubated on ice for 5 minutes. The lysed cells were then centrifuged for 5 minutes (13,000 x g at 4°C). The supernatant was added to 650 µl of isopropanol to precipitate the DNA. The sample was vortexed and then centrifuged for 10 minutes (13,000 x g at 4°C). The supernatant was carefully removed and the pellet was washed with 700 µl 70% EtOH. The DNA was centrifuged for a further 10 minutes (13,000 x g at 4°C) and the supernatant was removed. The pellet was dried in a vacuum centrifuge for 15 minutes. The dry pellet was resuspended in 200 µl ddH_2O. 5 µl of 10 mg/ml RNase A was added before the DNA was used for PCR.

2.2.4 Polymerase usage sequencing (Pu-Seq)

2.2.4.1 Strain growth, DNA preparation and Illumina library preparation

2.2.4.1.1 Cell collection and DNA extraction

Single colonies of *rnh201*Δ *cdc20-M630F* (Polε) and *rnh201*Δ *cdc6-L591G* (Polδ) Pu-Seq strains were each used to inoculate 10 mL YE. The cultures were grown overnight to saturation at 30°C. $1.8x10^7$ cells from the saturated primary culture were used to inoculate 1 L of YE. Cells were grown for ~17 hours to mid-log phase (concertation between $3x10^5$ and $5x10^5$ cells/mL). Doubling times of Pu-Seq strains with additional mutations varied. The amount of primary culture and/or growth times were changed accordingly.

800 mL of the secondary culture was pelleted and the cells were washed in 40 mL of ddH_20. The cells were resuspended in 5 mg/mL lyticase (Merrick and Fisher) in NIB buffer. The cells were incubated for 30 minutes at 37°C. The lysed cells were pelleted, washed in 20 mL ddH_2O and resuspended in 2 mL of Qiagen G2 digestion buffer. 100 µl of 10 mg/mL RNase A was added and the sample was incubated for 30 minutes at 37°C. 100 µl of 30% (w/v) N-lauroyl sarcosine and 100 µl of freshly prepared 20 mg/ml Proteinase K were added the sample was incubated for a further 30 minutes at 55°C.

The sample was centrifuged (4,000 x g at 4°C) for 15 minutes. The supernatant (S1) was removed and stored on ice and the pellet was resuspended in 1 mL

Qiagen G2 buffer. 50 µl of 30% (w/v) N-lauroyl sarcosine and 50 µl of freshly prepared 20 mg/ml Proteinase K were added to the resuspended pellet. The sample was incubated for a further 30 minutes at 55°C.

The resuspended pellet was centrifuged (4,000 x g at 4°C) for 15 minutes. The supernatant (S2) was added to S1 and the pellet was discarded. The pooled supernatant was added to Qiagen 100/G Genomic-tip column (equilibrated using 4 ml of Qiagen buffer QBT). The column was washed using 15 mL of Qiagen buffer QC. The DNA was eluted in 5 mL of Qiagen buffer QF (warmed to 55°C).

The 5 mL elute was split equally between five 2 mL microfuge tubes. 700 µl of isopropanol was added to each tube. The tubes were vortexed and then centrifuged for 15 minutes (13,000 x g at 4°C). The supernatant was carefully removed and the pellets were washed with 700 µl 70% EtOH. The DNA was centrifuged for a further 10 minutes (13,000 x g at 4°C) and the supernatant was removed. The pellets were dried in a vacuum centrifuge for 15 minutes. The dry pellets were resuspended in a total volume of 200 µl ddH_2O. The DNA concentration was quantified using a Nanodrop spectrophotometer.

2.2.4.1.2 Alkali treatment and size selection

20 µg of genomic DNA was incubated with 0.3 M NaOH for 2 hours at 55°C (recommended final volume 100 µl). Half of the reaction (10 µg of fragmented single stranded DNA) was run on 2% (w/v) TBE agarose gel in 0.5X TBE for 2 hours at 100 V. The gel was stained for 1 hour with 0.5 µg/ml acridine orange (Life Technologies) solution (2,000X dilution in water from 10 mg/ml stock). The gel was destained in 300 mL of ultra-pure water for 1 hour and then overnight in 300 mL of fresh ultra-pure water.

The gel was visualized under long-wave UV illumination. The fragments containing 300-500 bp single stranded DNA (ssDNA) were excised using a scalpel. The DNA was extracted from the agarose gel using Macherey-Nagel Gel and PCR Clean-up Kit, following manufacturer's instructions. The concentration of ssDNA was quantified using a NanoDrop® spectrophotometer.

2.2.4.1.3 Second (complementary) strand synthesis

100 ng of ssDNA was taken up to a final volume of 30 µl in ultra-pure water. The following reaction was set up in a clean PCR tube:

Component	Volume
ssDNA in ultra-pure water	30 µl
8N random primers (3 mg/mL)	5 µl
NEB 2.1 buffer	5 µl

The reaction was incubated at 95°C for 5 minutes in a thermocycler and then for 5 minutes on ice. Immediately afterwards, 5 µl of dNTPs (where dTTP was substituted for dUTP; 2 mM each) and 1 µl of T4 DNA polymerase were added. The reaction was incubated at 37°C for 20 minutes in a thermocycler. Following this, the reaction was quenched using 5 µl of pH 8.0 EDTA.

The 55 µl reaction containing the dsDNA fragments was transferred to a fresh microfuge tube. 99 µl (1.8X) of AMPure XP beads (Beckman Coulter) were added and left at room temperature for 5 minutes. The beads were separated from the reaction using a magnetic rack. The supernatant was removed and discarded. The beads were washed twice with 200 µl of fresh 80% EtOH. The beads were left to dry on the magnetic rack for 5 minutes before being resuspended in 55 µl of ultra-pure water. The reaction was left at room temperature for 5 minutes and the beads were separated from the supernatant using a magnetic rack. 53 µl was carefully taken up.

1 µl was used to assess the size of DNA fragments using a high-sensitivity DNA Bioanalyser chip (Agilent). 50 µl was used to prepare the Illumina library.

2.2.4.1.4 Illumina library preparation

The libraries were prepared using the NEBNext® Ultra™ II DNA Library Prep Kit (NEB). The manufacturer's protocol was followed with some modifications. The final protocol is described below.

End repair

Component	Volume
End Prep Enzyme Mix	3 µl
End Prep Reaction buffer	7 µl
dsDNA (from previous step)	50 µl

Reaction was incubated in a thermocycler (with lid heated to 98°C) using the following conditions:

20°C	30 minutes
65°C	30 minutes
4°C	Hold

Adaptor ligation

The following reaction was set up and incubated in a thermocycler (with lid heated to 30°C) for 20 minutes at 37°C.

Component	Volume
End repair reaction mixture	60 µl
Ligation Master Mix	30 µl
1/10 diluted NEBNext® adaptors	2.5 µl
Ligation enhancer	1 µl

This allowed the ligation of the hairpin loop-shaped NEBNext adaptors, which ligate in a known orientation relative to the 3' and 5' ends of the short DNA fragments. A directional library created in this manner contained fragments whose 3' and 5' ends are flanked by unique sequences.

The fragments with the ligated adaptors were size selected using AMPure XP beads (Beckman Coulter). The final volume of the reaction was adjusted to 100 µl using ultra-pure water in a clean microfuge tube. 35 µl (0.35X) of AMPure XP beads was added, mixed with the DNA and left at room temperature for 5 minutes. The beads were separated from the supernatant using a magnetic rack. The supernatant was carefully removed and added to a clean microfuge tube. 35 µl (0.26X) of AMPure

XP beads was added mixed with the DNA and left at room temperature for 5 minutes. The beads were separated from the supernatant using a magnetic rack. The supernatant was removed and discarded. The beads were washed three times with 200 µl of fresh 80% EtOH. The beads were left to dry on the magnetic rack for 5 minutes before being resuspended in 20 µl of ultra-pure water. The reaction was left at room temperature for 5 minutes and the beads were separated from the supernatant using a magnetic rack. 17 µl was carefully taken up.

USER digestion and PCR enrichement

17 µl of size selected dsDNA fragments with ligated adaptors were used to set up the following reaction:

Component	Volume
USER™ Enzyme	3 µl
Q5 Master Mix	25 µl
Universal primer (10 µM)	2.5 µl
Index primer (10 µM)	2.5 µl
dsDNA	17 µl

Reaction was incubated in a thermocycler (with lid heated to 98°C) using the following conditions:

37°C	15 minutes	
98°C	30 seconds	
98°C	10 seconds	X9
65°C	75 seconds	
65°C	5 minutes	
4°C	Hold	

50 µl (1X) of AMPure XP beads was added to the 50 µl reaction, mixed with the DNA and left at room temperature for 5 minutes. The beads were separated from the supernatant using a magnetic rack. The supernatant was removed and discarded. The beads were washed twice with 200 µl of fresh 80% EtOH and left to dry on the magnetic rack for 5 minutes before being resuspended in 26 µl of ultra-pure water. The reaction was left at room temperature for 5 minutes and the beads were separated from the supernatant using a magnetic rack. 25 µl was carefully taken up.

25 µl (1X) of AMPure XP beads was added to the 25 µl sample, mixed with the DNA and left at room temperature for 5 minutes. The beads were separated from the supernatant using a magnetic rack. The supernatant was removed and discarded. The beads were washed twice with 200 µl of fresh 80% EtOH and left to dry on the magnetic rack for 5 minutes before being resuspended in 23 µl of ultra-pure water. The reaction was left at room temperature for 5 minutes and the beads were separated from the supernatant using a magnetic rack. 20 µl was carefully taken up.

1 µl of a 1/5 dilution of the final library was run on the Agilent 2100 Bioanalyser, using a high-sensitivity DNA chip. The rest was stored at -20°C.

2.2.4.2 Illumina library sequencing

Between 28 and 30 Pu-Seq libraries were multiplexed and sequenced using NextSeq 550. The runs typically generated 400 million paired end reads, 101 bp long. Each library, therefore, generated ~ 13 million reads.

2.2.4.3 Data analysis

The Pu-Seq data analysis can be broadly subdivided into two parts:

1) The analysis of the raw reads. Briefly, this analysis included the alignment of reads and tracking of the 5' end of the reads, counting the number of times they appeared in each bin.
2) Analysis of the count data to produce tracks of polymerase usage, from which origins of replication, local replication timing and polymerase bias were calculated.

Both parts of the analysis are described in detail below. Flowcharts describing the steps for parts 1) and 2) are also shown in Figs. 2.1 A and B, respectively.

2.2.4.3.1 Analysis of the raw reads

Data from paired end sequencing of each library resulted in two FASTQ files, each with R1 or R2 sequencing data. The reads in each file were aligned to the *S. pombe* reference genome SP2 using Bowtie2. In addition to aligning the reads, Bowtie2 also trimmed 1 bp off the 5' end of the read (--trim5 1) and 30 bp off the 3' end of the read (--trim3 30). The paired end alignment resulted in a single SAM file for each strain.

Each SAM file was processed using the pe-sam-to-bincount.pl program (https://github.com/yasukasu/sam-to-bincount), written by Dr. Yasukazu Daigaku. The program analyses the data in the following way:

- Concordantly aligned reads (i.e., reads where both of the paired end reads were aligned) are selected from the SAM file (by filtering the alignment records using the 0x2 flag bit) and sorted.
- The genome is divided into bins (default bin size is 300 bp).
- The aligned reads are sorted depending on whether they originated from the forward or reverse strand. This was carried out using the directionality information that was preserved during library construction (as described in 2.2.4.1.4).
- The position of the base directly next to R1-end (compensating for the 1 bp 5' trimming carried out by Bowtie2) is assigned to each bin.
- The number of times the 5' end was found in each bin is counted and returned in CSV format.

The SAM file analysis produces two CSV files for each library, each containing counts for the Watson and Crick strands. The analysis is carried out for the reads from the polymerase ε and δ libraries and the four CSV files are used in the next step of the analysis.

2.2.4.3.2 Analysis of the count data

The raw count data in the CSV files was further analysed using a custom Pu-Seq R script (Appendix 9.1), written by Dr. Andrea Keszthelyi. The program analyses the data in the following way:

- Count data for the Watson and Crick strands, from the polymerase ε and δ experiments are input (denoted $C_W{}^{\delta}$, $C_W{}^{\varepsilon}$, $C_C{}^{\delta}$, $C_C{}^{\varepsilon}$, respectively).
- The count data for each dataset in each bin (x) are normalised (N) to the total number of counts in each data set, for example :

$$N_{W(x)}{}^{\delta} = C_{W(x)}{}^{\delta} / \sum (C_{W(x)}{}^{\delta})$$

- Assuming that each strand can be synthesised by only polymerase ε or δ, the contribution of each polymerase to the synthesis of each strand was

calculated at each bin. For example, to calculate the ratio of polymerase δ usage on the Watson strand ($R_W{}^{\delta}$):

$$R_{W(x)}{}^{\delta} = N_{W(x)}{}^{\delta} / (N_{W(x)}{}^{\delta} + N_{W(x)}{}^{\varepsilon})$$

- The data were smoothed using a moving average 3, i.e., the data point for each bin (x) is an average of seven bins (bin at position x, three bins upstream and three bins downstream).

The usage of the polymerases at each strand was further used to extract the bias of polymerase usage at every position, positions and the firing efficiencies of the origins of replication, the progression of the leftward moving replication forks and the local replication timing. The details of these calculations are described below.

<u>Calculating the bias of polymerase usage on both strands</u>

When calculating the usage of each polymerase on the Watson and Crick strands (i.e., the ratios of polymerase usage), it is assumed that only polymerases ε and δ replicate the DNA. The contribution of each polymerase at different positions, however, may differ and can be calculated.

Total polymerase usage on both strands can be counted for each polymerase:

$$R_{CW(x)}{}^{\delta} = R_{C(x)}{}^{\delta} + R_{W(x)}{}^{\delta}$$

Given that the sum of polymerase δ and ε usage at every position is equal to 1, i.e.:

$$R_{C(x)}{}^{\delta} = 1 - R_{C(x)}{}^{\varepsilon}$$

$$R_{W(x)}{}^{\varepsilon} = 1 - R_{W(x)}{}^{\delta}$$

Therefore, the sum of the totals must be equal to 2, i.e.:

$$R_{(x)}{}^{POL} = R_{(x)}{}^{\delta} + R_{(x)}{}^{\varepsilon} = 2$$

From this, the bias towards the relative contribution of polymerase δ on both strands (B) can be counted using:

$$B_{(x)}{}^{\delta} = (R_{C(x)}{}^{\delta} + R_{W(x)}{}^{\delta}) / 2$$

The expected contribution of each polymerase at every position (x) is 0.5. Deviation from $B_{(x)}{}^{\delta}$ above and below 0.5 is therefore considered a bias towards the usage polymerases δ and ε, respectively.

The same calculation can be done for polymerase ε. In the work carried out for this thesis, however, $B_{(x)}{}^{\delta}$ was used to calculate polymerase biases.

Calculating the positions of the origins of replication and their respective firing efficiencies

Given that the sum of polymerase δ and ε usage at every position is equal to one, i.e.:

$$R_{C(x)}{}^{\delta} = 1 - R_{C(x)}{}^{\varepsilon}$$

$$R_{W(x)}{}^{\varepsilon} = 1 - R_{W(x)}{}^{\delta}$$

The usage of polymerase δ on the Crick (reverse) strand ($R_C{}^{\delta}$) and polymerase ε on the Watson (forward) strand ($R_W{}^{\varepsilon}$) contain all of the count information. Origin positions and efficiencies were calculated using the steep transitions in the $R_C{}^{\delta}$ and $R_W{}^{\varepsilon}$ datasets separately (marked with yellow circles on Fig. 2.1 B ii).

To determine which of the transitions would be classed as origins, differentials of each data set (smoothed using a moving average of 3, as described for the data normalisation) were calculated. In order to reduce noise, only positive peaks in the differential data that had a value above the set threshold (marked with a green line on Fig. 2.2.2B ii) were further analysed. The default threshold was set to the 30th percentile of all peaks. Given the stochastic nature of origin firing in *S. pombe*, peaks within 4 bins were merged and considered as one "origin of replication". The midpoint position of the peaks merged was assigned as the position of the origin.

The efficiency of each origin was calculated from the difference between the maxima and the minima of the steep transition zone (marked with yellow triangles on Fig. 2.1 B ii), seen in the original $R_C{}^{\delta}$ and $R_W{}^{\varepsilon}$ datasets. The difference, multiplied by 100 is the probability of the origin firing in a population of cells as a percentage.

Each of the positions of the origins and their corresponding efficiencies calculated from the Watson and Crick datasets were then averaged and a final list of the positions of origins of replication and their respective origin firing efficiencies was produced.

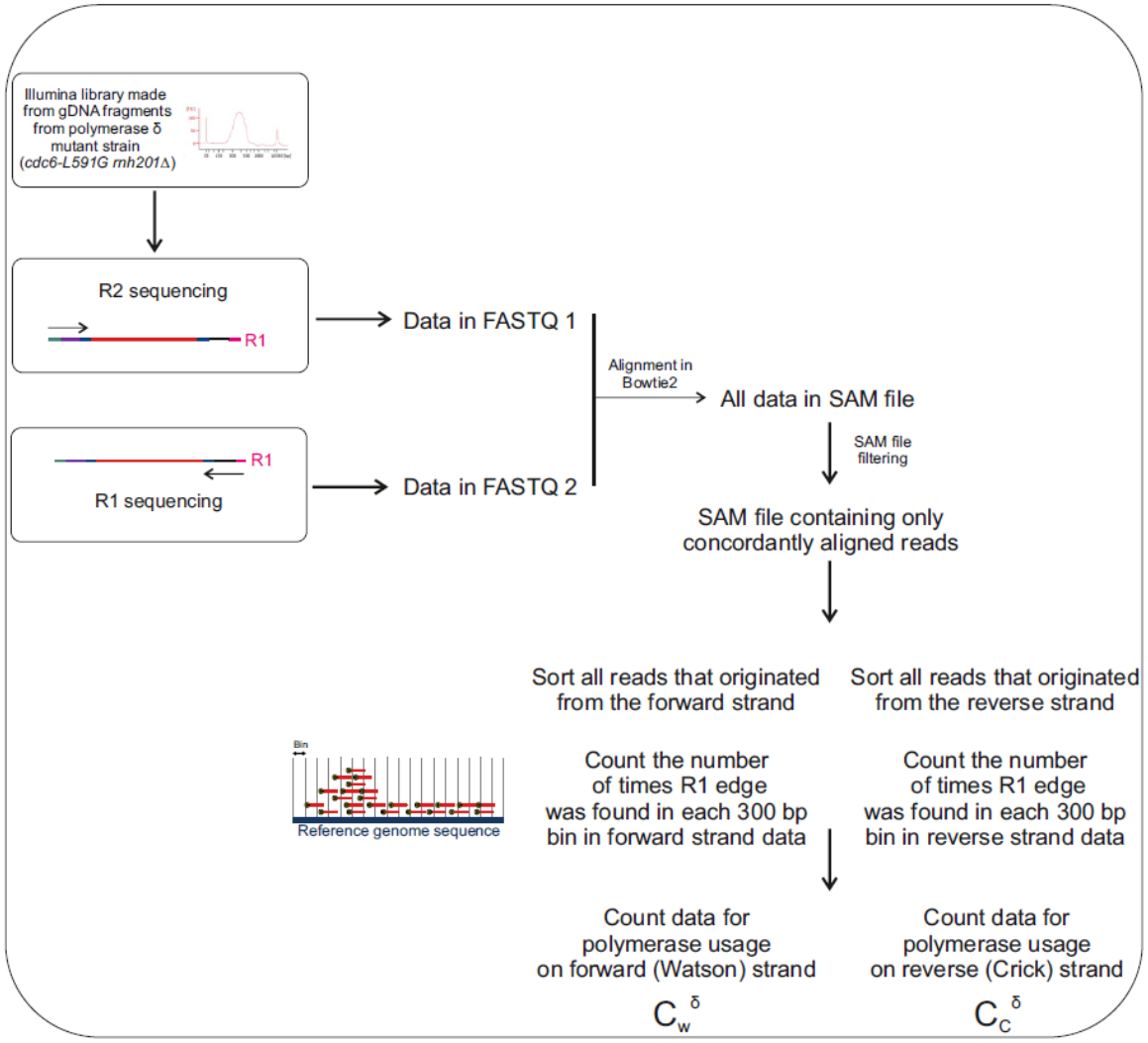
Calculating the progression of leftward moving forks and relative replication timing

The leftward moving forks on the 5' to 3' forward (Watson) strand are synthesised using polymerase δ. Conversely, on the Crick strand, the leftward moving forks are synthesised using polymerase ε. As such, the average progression of leftward moving forks at each position (x) was calculated using:

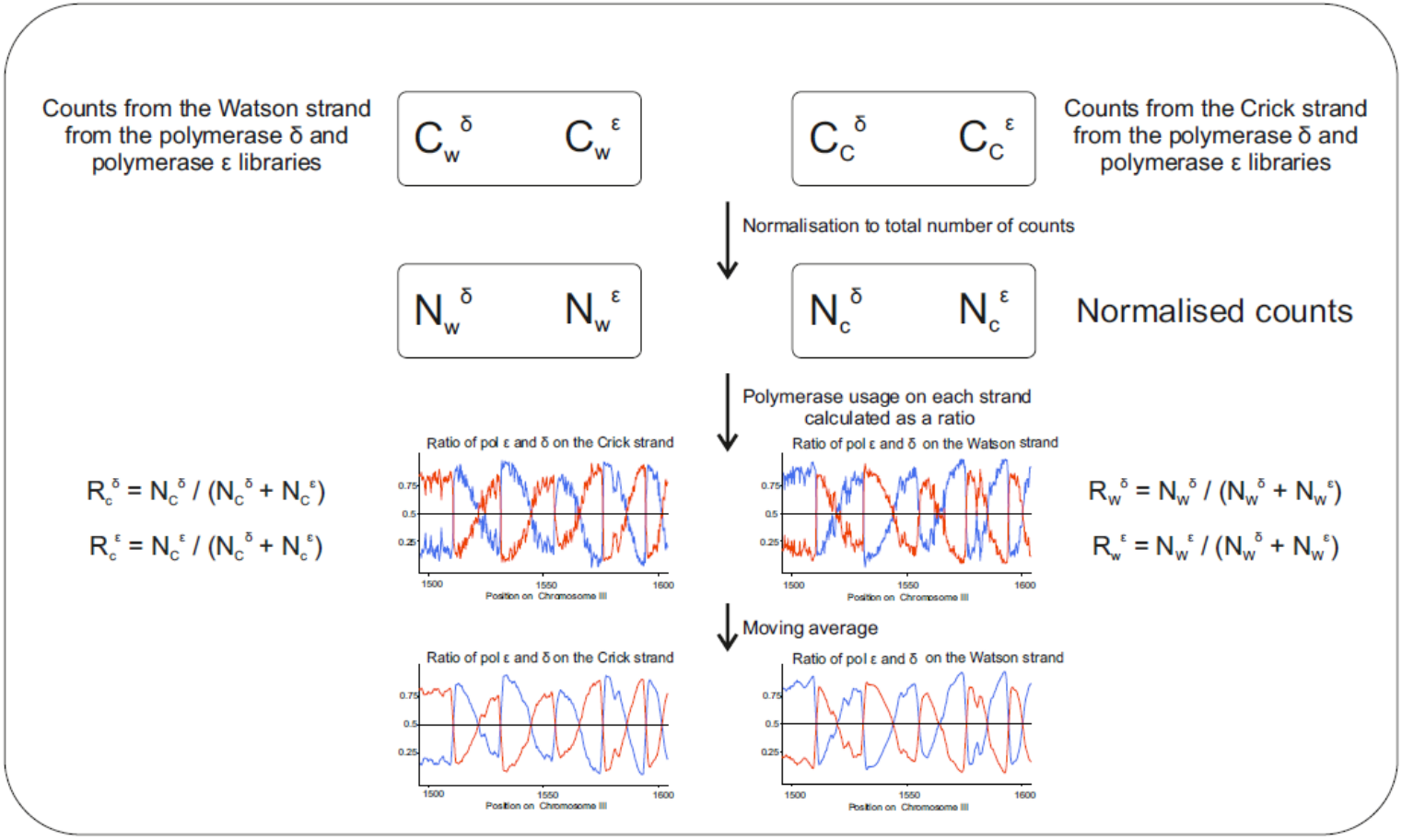
$$\frac{{R_W}^{\delta} + {R_C}^{\varepsilon}}{2}$$

Using the average progression of leftward moving forks and assuming a mean replication fork velocity of 1.5 kb/min, the relative replication timing (Pu-Seq Trep) was calculated for each bin, as described by (Retkute, Nieduszynski et al. 2012). Trep calculations can be found in the Pu-Seq R script (9.1).

A) Analysis of raw reads



Bi) Analysis of count data - polymerase usage



Bii) Analysis of count data - Origins of replication

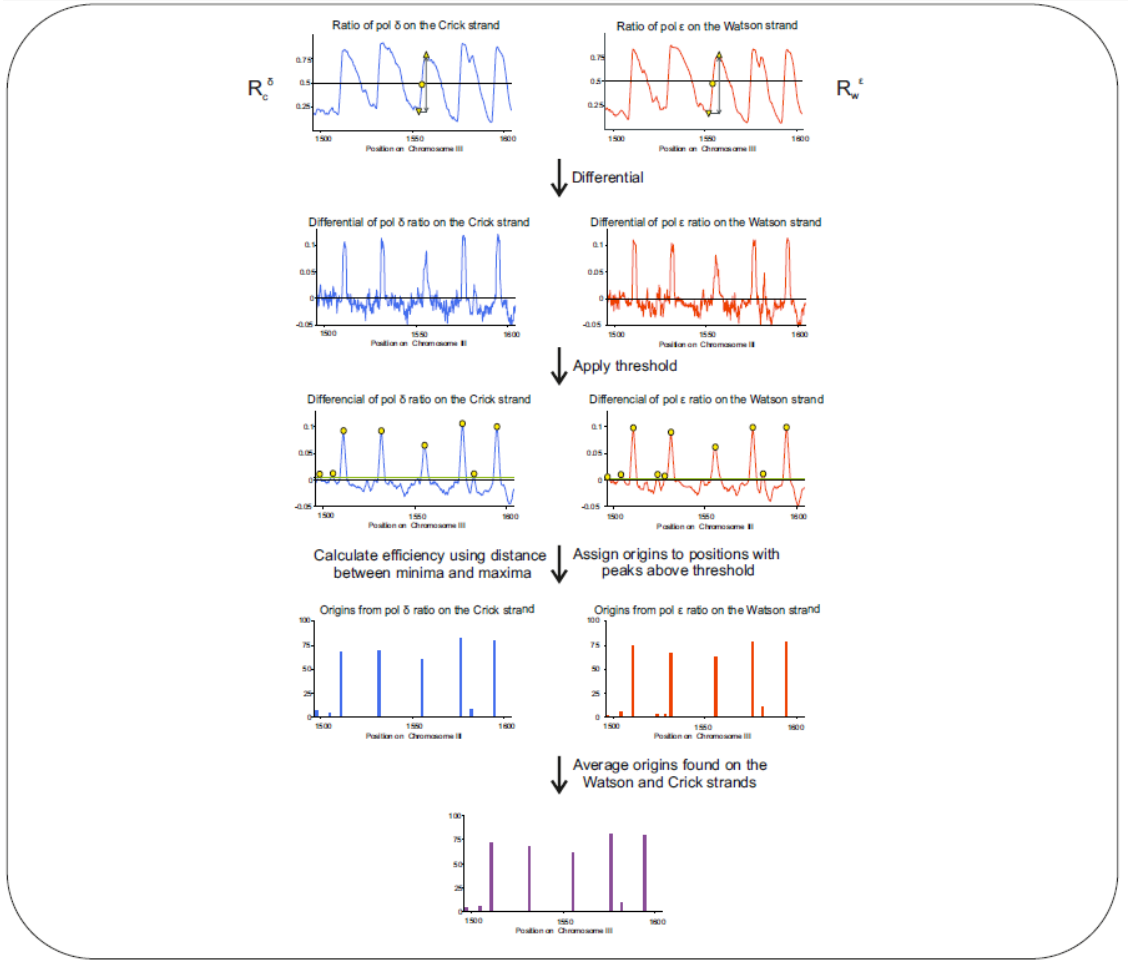


Fig. 2.1 - **Steps involved in Pu-Seq data analysis**

A) Analysis of raw reads. Each Pu-Seq library is sequenced using Illumina paired end sequencing. The data sequenced using the forward and reverse primers are stored in FASTQ format - files FASTQ1 and FASTQ2, respectively. The data are then aligned to the SP2 reference genome using Bowtie2, set to end-to-end alignment in mixed mode. Concordantly aligned reads are extracted and the number of reads in each 300 bp bin is counted across the genome. This analysis is carried out for reads generated by Pu-Seq polymerase ε and δ libraries separately (the flowchart shows the analysis for one of the libraries). Raw counts of polymerase usage on the forward (Watson) and reverse (Crick) strands are produced.

B) Analysis of raw counts.

i) Calculating the polymerase usage ratios. The counts of polymerase usage are normalised for each dataset separately. Polymerase usage of each polymerase on the forward (Watson) and reverse (Crick) strands is calculated, using the equations shown on the figure. The resultant data are smoothed using a moving average.

ii) Defining the position of the origins of replication and calculating their respective firing efficiencies. The differentials of the polymerase usage of polymerases δ and ε on the forward (Watson) and reverse (Crick) strands, respectively, are calculated. The differential peaks > threshold are marked with yellow circles on panel 3. These positions are identified on the ratios of polymerase usage (panel 1). The distance between the minimum and maximum values (yellow triangles, panel 1) is measured to represent the efficiency of firing of that origin. The origin efficiencies calculated from the usage of polymerases δ and ε are averaged to produce the final efficiency of firing of each origin of replication.

Adapted from (Keszthelyi, Daigaku et al. 2015).

2.2.5 Elutri-Seq

2.2.5.1 Cell collection, DNA and library preparation and sequencing

2.2.5.1.1 Cell collection and synchronisation

100 mL YEP was inoculated and grown overnight at 30°C. The next morning, the concentration of cells was counted using a haemocytometer. The cells were diluted to $7.5x10^5$ cells/mL (i.e., between two or three doublings away from mid-log phase).

The mid-log phase sample was used to inoculate $7.8x10^6$ cells into 1 L YEP. Four 1 L YEP cultures were set up and grown for 8 doublings (20 hours). 4 L of cells between $2x10^6$ and $4x10^6$ cells/mL were used for elutriation. $2.5x10^7$ G_2 cells were collected and resuspended in YE media to a final concentration of $2.5x10^5$ cells/mL. The cells were allowed to complete one cell cycle at 30°C.

5 mL samples were taken at 20 minutes and then every 10 minutes between 50 and 120 minutes after release into YE media, for DNA extraction. The cells were collected into 0.05 mM sodium azide and stored on ice until the end of the time course. Cells were then pelleted, washed with ddH_2O and snap frozen in liquid nitrogen.

2.2.5.1.2 Analyzing cells cycle progression

The cell cycle was monitored using flow cytometry (FACS) and microscopy, as described below.

For microscopy analysis, 1 mL samples were taken every 10 minutes between 20 and 120 minutes after release into YE media. The cells were pelleted and resuspended in 1 mL methanol. 3 µl of the cells were mounted onto a glass slide, allowed to dry, and stained with 1 µg/mL 4′,6′-diamidino-2-phenylindole (DAPI) and 2.5% v/v calcofluor-white. At least 100 cells were counted at each time point, using an inverted fluorescence microscope (EVOS™ FL). Based on morphology, the stage at which each cell was in the cell cycle was scored.

For FACS analysis, 5 mL samples were taken at 20 minutes and then every 10 minutes between 50 and 120 minutes after release into YE media. The cells were collected into 200 mM ice-cold EDTA and 0.05 mM sodium azide and stored on ice

until the end of the time course. Cells were then pelleted, washed with ddH_2O, resuspended in 1 mL 70% EtOH and stored at 4°C.

To prepare cells for FACS, the 5 mL aliquots were pelleted, washed and resuspended in 500 µl of 50 mM trisodium citrate (pH 7.0) containing 50 µl of 10 mg/mL RNase. Cells were incubated for 3 hours at 37°C. A master mix containing 10 µl of 500 µg/mL propidium iodide (PI) (Sigma) and 1 mL of 50 mM tri-sodium citrate (pH 7.0) was prepared, per sample. 200 µl of the RNase treated cells was added to 1mL of the master mix and the cells were then sonicated for 10 seconds at 20% power (Ultra sonic Processor sonicator). Cells were analysed for DNA content on BD Accuri™ C6 Plus flow cytometer.

2.2.5.1.3 Genomic DNA extraction, sonication, library preparation and sequencing

Genomic DNA was extracted for the S-phase and G_2 time points as follows. The cells were thawed and resuspended in 1 mg/mL lyticase (Merrick and Fisher) in CSE and incubated at 37°C for 30 minutes. The cells were pelleted again and resuspended in 450 µl of 5X TE. 50 µl of 10% SDS was added and the sample was incubated at room temperature for 5 minutes. 150 µl of 5 M KAc was added and the samples were incubated on ice for 5 minutes. The lysed cells were centrifuged down for 5 minutes (13,000 x g at 4°C). The supernatant was added to 650 µl of isopropanol to precipitate the DNA, vortexed and centrifuged for 10 minutes (13,000 x g at 4°C). The supernatant was carefully removed and the pellet was washed with 700 µl 70% EtOH. The DNA was centrifuged for a further 10 minutes (13,000 x g at 4°C) and the supernatant was removed. The pellet was dried in a vacuum centrifuge for 15 minutes. The dry pellet was resuspended in 250 µl ultra-pure water.

5 µl of 10 mg/mL RNase was added to the DNA and incubated for 20 minutes at 37°C. 2 µl of 10% SDS and 10 µl of 10 mg/mL Proteinase K was added and the DNA was incubated for 1 hour at 55°C. The volume was taken up to 500 µl using ultra-pure water. 2 volumes (1 mL) of phenol:chloroform:isoamyl alcohol (25:24:1, Sigma, 77617) was added, the sample was vortexed for 20 seconds and centrifuged for 5 minutes (13,000 x g at 4°C). The upper phase (containing the DNA) was transferred to a fresh microfuge tube. The DNA was precipitated by adding 1/10 volume of 3 M NaAc, 2.5 volumes of 100% EtOH and incubating on ice

for 10 minutes. The sample was centrifuged for 15 minutes (13,000 x g at 4°C), the supernatant carefully removed and the pellet washed in 1 mL 70% EtOH. The DNA was centrifuged for a further 10 minutes (13,000 x g at 4°C) and the supernatant was removed. The pellet was dried in a vacuum centrifuge for 15 minutes. The dry pellet was resuspended in 100 µl ultra-pure water and transferred to a Covaris microTUBE.

The DNA was sonicated for 6 minutes and 50 seconds using the Covaris M220 system (duty cycle 20%, peak incident power 50 Watts, 200 cycles per burst, water bath temperature 20°C). The sonicated DNA was size selected to enrich for ~250 bp fragments using AMPure XP beads (Beckman Coulter). 90 µl (0.9X) of AMPure XP beads was added, mixed with the DNA and left at room temperature for 5 minutes. The beads were separated from the supernatant using a magnetic rack. The supernatant was carefully removed and added to a clean microfuge tube. 20 µl (1.1X) of AMPure XP beads was added mixed with the DNA and left at room temperature for 5 minutes. The beads were separated from the supernatant using a magnetic rack. The supernatant was removed and discarded. The beads were washed twice with 200 µl of fresh 80% EtOH and left to dry on the magnetic rack for 5 minutes before being resuspended in 55 µl of ultra-pure water. The reaction was left at room temperature for 5 minutes and the beads were separated from the supernatant using a magnetic rack. 53 µl was carefully taken up.

1 µl was run on the Agilent 2100 Bioanalyser, using a high-sensitivity DNA chip.

50 µl was used to construct an Illumina library which was prepared and sequenced, as described in 2.2.4.1.4 The depth of sequencing is specified for each Elutri-Seq experiment individually.

2.2.5.2 Data analysis

Elutri-Seq RT profiles were generated by calculating the ratio between the number of reads that map to each locus (i.e., the copy number) in early S-phase and G_2. Loci that replicated earlier had a greater copy number than those that replicated late in S-phase.

The FASTQ data were aligned to the SP2 reference genome using Bowtie2, as described in 2.2.4.3. In order to calculate these ratios, the following samtools and

bedtools commands were used (kindly shared by Prof. Conrad Nieduszyski) on the S-phase (FILE1) and G_2 (FILE2) data (file names are shown in bold for clarity):

1) Sort and index reads

```
samtools sort FILE1.bam FILE1_sorted

samtools index -b FILE1_sorted.bam
```

2) Retrieve and print the sequence names in the index file.

```
samtools idxstats FILE1_sorted.bam | awk 'BEGIN {OFS="\t"} {if ($2>0) print ($1,$2)}' > FILE1_genome
```

3) Bin the genome into 1 kb windows

```
bedtools makewindows -g FILE1_genome -w 1000 > windows.bed
```

4) Determine the number of 5' ends of reads from control sample that map to each position in the genome

```
samtools view -h -@ 1 -q 30 -F 3840 -f 64 -L windows.bed FILE1_sorted.bam | grep -v XS:i: | samtools view -b - | bedtools genomecov -5 -d -ibam stdin | awk 'BEGIN {OFS="\t"} {if ($3>0) print $1,$2,$2,"GENOTYPE",$3}' > FILE1.bed
```

5) Sum reads in each bin and convert to bed format (done in awk).

```
bedtools map -a windows.bed -b FILE1.bed -null 0 -o sum | awk  'BEGIN {OFS="\t"} {if ($4>0) print $1,$2,$3,"GENOTYPE",$4} > FILE1_total_reads_in_bins.bed
```

6) Normalise the number of reads in each bin to the total number of reads (done in R).

```
File1<- read.delim("FILE1_total_reads_in_bins.bed", header=FALSE, stringsAsFactors=FALSE)

File1$V5<-(File1[,5]/sum(File1[,5]))

write.table(File1, "FILE1_total_reads_in_bins_normalised.bed", quote = F, col.names = F, row.names = F, sep="\t")
```

7) Repeat steps 4-6 for FILE2.
8) Use bedtools to calculate the ratio between the two files. Windows with less than a quarter of the expected number of reads mapping were excluded.

```
bedtools intersect -a FILE1_total_reads_in_bins_normalised.bed -b
FILE2_total_reads_in_bins_normalised.bed -wb | awk 'BEGIN {OFS="\t"}
{line[NR] = $0; control+=$5; repl+=$10; count+=1} END { for (r=1;
r<=NR; ++r) {split(line[r], fields); if(fields[5]>0.25*control/count
&& fields[10]>0.25*repl/count) print
fields[1],fields[2],fields[3],fields[4],(fields[10]/fields[5])}}' >
ElutriSeq_ratio.bed
```

2.2.5.3 Calculating the percentage of genome replicated from FACS and septation data

The method for calculating the percentage of the genome replicated was kindly carried out and shared by Dr. Carolin Müller from Prof. Nieduszynski's lab.

During the preparation of cells for FACS, the post S-phase *S. pombe* cells separate. As a result, the 2N peak can be assigned to post-S-phase G_2 cells (2N), G_2 cells (2N), M-phase cells (2N) and G_1 cells (2 x 1N).

The proportion of cells post-S-phase were determined using cell morphology for each time point, described in 2.2.5.1.2. This was low for early time points (e.g., 20-60 minutes after elutriation). The proportion of S-phase cells increased during the time course with the onset and completion of S-phase.

The fraction of post-S-phase cells (morphology data) were then used to calculate how much of the 2N peak (FACS data) could be explained by cells that have completed S-phase. Using this, the population average of the percentage of the genome replicated was calculated at each time point.

2.2.6 Imaging *S. pombe* with mEos3.2 and GFP tagged proteins

Live *S. pombe* with mEos3.2 tagged proteins were imaged with a custom-built microscope similar to that described in (Etheridge, Boulineau et al. 2014). Briefly, the microscope was built around an Olympus IX73 inverted microscope body fitted with a motorized stage (Prior H117E1I4), 60x objective (Olympus APON-TIRF 1.45NA) and a 30^{o}C heated chamber (Digital Pixel Ltd). Excitation of fluorescent proteins was achieved with a trio of lasers: 561 nm (Cobolt, Jive), 488 nm (Toptica, iBeam) and a 405 nm lasers (LaserBoxx, Oxxius). The path of each laser beam was expanded and collimated, and entry into the microscope was controlled by individual automated shutters (Prior Scientific). A multiband dichroic (405

nm/488 nm/561nm/635-25 nm) was used to reflect laser light into the sample and the resulting emission from fluorophores was filtered using either a 525-40 nm or a 593-40 nm (Semrock) bandpass filter for GFP and mEos3.2 respectively. Emission fluorescence was directed out of the microscope onto an EMCCD camera (Photometrics Evolve Delta). The image was expanded prior to the camera using a 2.5x photo-eyepiece to achieve a final image pixel size of 104 nm.
The microscope was automated using custom written scripts in Micro-Manager 1.4. After focussing into the mid-plane of the sample the sequence of data acquisition was as follows. Gar2-GFP was excited using the 488 nm laser at 15% power and images were acquired using 100ms exposure time. Rif1-mEos3.2 and Rif1PP11-mEos3.2 were excited using dual continuous 405 nm and 561 nm excitation (1 W/cm^2 and 1 kW/cm^2 respectively) for 1000 frames with 50 ms exposure time. Multiple fields of view were acquired per experimental repeat.

Raw mEos3.2 single molecule data was processed using the GDSC SMLM software package as described in (Etheridge, Boulineau et al. 2014).

Chapter 3

3.1 Introduction

Genome wide mapping of ϒH2A suggested that repetitive and constitutively heterochromatic regions may act as RFBs in *S. pombe* (Rozenzhak, Mejia-Ramirez et al. 2010). The levels of ϒH2A moderately increased at these sites during unperturbed S-phase, suggesting that the process of heterochromatin formation or the final chromatin structure could impede fork progression. As discussed in 1.3.3, chromatin remodellers disassemble chromatin in order to allow replication and transcription facttors to access naked DNA (Demeret, Vassetzky et al. 2001). While it is possible that not all of the nucleosomes unpack in time and act as endogenous RFBs, more studies are necessary to establish a definite link between heterochromatin and replication fork stalling.

As discussed in 1.2.3.2, when the burden of replication stress is high, the intra-S-phase checkpoint is activated to inhibit origin firing and stall replication forks (Lambert and Carr 2005). Collapsed replication forks can be restarted during S-phase using an HR dependent and DSB independent strand invasion mechanism (Mizuno, Lambert et al. 2009, Lambert, Mizuno et al. 2010). HR-restarted replication forks are not only more error prone (Iraqui, Chekkal et al. 2012) but also not canonical (Miyabe, Mizuno et al. 2015). At these forks, both the leading and lagging strand replication are carried out by polymerase δ (Miyabe, Mizuno et al. 2015). It has, therefore, been proposed that a local bias towards polymerase δ usage can act as a marker of fragile sites.

Polymerase usage sequencing (Pu-Seq) is a deep-sequencing method developed in *S. pombe* to track the global usage of polymerases ε and δ on both the forward and reverse strands (Daigaku, Keszthelyi et al. 2015). From these data, a bias towards the usage of each polymerase genome wide can be calculated (calculations described in 2.2.4.3.2). So far, it has not been possible, however, to accurately track polymerase usage at repetitive regions. This is due to a limitation of the method - the poor coverage of repetitive sequences by short reads.

We have, therefore, set out to optimise the Pu-Seq method for repetitive regions, to allow a more accurate assessment of polymerase usage and polymerase bias at these loci.

3.1.1 Repetitive regions in *S. pombe*

There are four repetitive and constitutively heterochromatic loci in *S. pombe* - centromeres, telomeres, the *mat* locus and the rDNA. All four are associated with *dg* and *dh* repeats, whose transcription initiates the RNAi mediated mechanism of heterochromatin formation (Cam, Sugiyama et al. 2005) (described in 1.3.2; Fig. 1.3).

Using any of these four loci as a representative region to study polymerase usage across repetitive DNA has a number of advantages and disadvantages. The *mat* locus contains a 7.5 kb long repetitive region (*cenH*), which shares 96% sequence similarity with the *dg* and *dh* repeats from Centromere 2 (Grewal and Klar 1997). The repetitive regions at telomeres, centromeres and rDNA all span longer regions (Cam, Sugiyama et al. 2005), which would allow polymerase usage to be observed over longer distances. The SP2 reference genome, however, also does not contain the *mat cenH* or any telomere and rDNA sequences and the centromeres are not fully annotated.

Despite their relatively poor annotation in the reference genome, a lot is known about the organisation of *S. pombe* centromeres. Although they differ in size, all three centromeres are composed of three types of elements organised in the same fashion- the outer (*otr*) region (which contains the *dg* and *dh* repeats) flanks the imperfect inverted inner (*imr*) repeats and the central region (*cnt*). The size of centromeres is dictated by the number of repeats present in the *otr* region. The shortest centromere, Centromere 1 (schematic shown in Fig. 3.1), has only one copy of each of the repeats (Wood, Gwilliam et al. 2002).

The *cnt* regions have over 48% similarity, over a 1,405-bp region, between the three centromeres. The *otr* region contains the 4.4 kb *dg*, 4.8 kb *dh* and 0.3 kb *cen253* repeats. The *dg* repeats are the most conserved - they share over 97% identity, over a 1,405-bp region (Wood, Gwilliam et al. 2002). The conservation of repeats between chromosomes poses a problem during alignment. Due to the

heuristic nature of the alignment algorithms, reads originating from the repeats could be mapped randomly between the chromosomes, if at all.

Strong conservation, however, is only seen between the *dg* repeats and only two of those repeats are present on Centromere 1. The structure of Centromere 1 and the sequence of the repeats in the SP2 reference genome correlate well with data from the early Southern Blotting and Sanger sequencing experiments (Nakaseko, Adachi et al. 1986, Nakaseko, Kinoshita et al. 1987, Takahashi, Murakami et al. 1992). We have, therefore, decided to optimise the Pu-Seq protocol to increasing the coverage of Centromere 1 and use it as a representative region of repetitive and heterochromatic DNA.

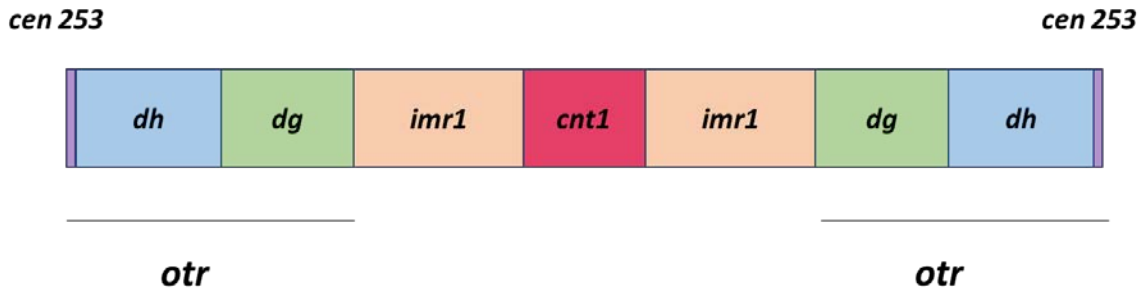


Fig. 3.1 - **Schematic representation of Centromere 1**

The central region (*cnt*) is flanked by the imperfect inverted repeats (*imr*). The *otr* repeats are composed of the *dg*, *dh* and *cen 253* repeats. Adapted from (Wood, Gwilliam et al. 2002).

3.1.2 Overview of the Pu-Seq protocol

As discussed in 1.2.2, at eukaryotic replication forks the leading and lagging strands are replicated by polymerases ε and δ, respectively (Nick McElhinny, Gordenin et al. 2008). Pu-Seq utilizes strains with mutations in the steric gates of the replicative polymerases ε (*cdc20-M630F*) and δ (*cdc6-L591G*) (Daigaku, Keszthelyi et al. 2015). These mutations allow the incorporation of a greater than average number of ribonucleotides into DNA. As a result, strains carrying the Pu-Seq mutation in polymerase δ "mark" their lagging strands with a heavy load of ribonucleotides. The same is true for the leading strands replicated in strains carrying the Pu-Seq mutation in polymerase ε. Ribonucleotides erroneously incorporated under wild type conditions are removed by the ribonucleotide excision repair pathway, which is initiated by the activity of RNase H2 (Sparks, Chon et al. 2012). To avoid ribonucleotide excision, the Pu-Seq strains carry an additional deletion of the catalytic subunit of RNase H2 - *rnh201*Δ.

A flowchart summarising the Pu-Seq protocol is shown in Fig. 3.2. Genomic DNA is extracted from Pu-Seq strains (*cdc20-M630F rnh201*Δ and *cdc6-L591G rnh201*Δ) and fragmented at the position of ribonucleotide incorporation using alkaline treatment. The fragmentation is caused by a hydroxyl attack on the 2'OH group of the ribose sugar (Fig. 3.3).

The protocol includes an optional step of size selecting small (300-500 bp) ssDNA fragments, before continuing with the complementary strand synthesis (Keszthelyi, Daigaku et al. 2015). The size selection is carried out by separating the ssDNA on a 2% agarose gel and excising the fragments of interest. The step is optional because dsDNA fragments are size selected during the library preparation steps. Despite this downstream size selection for smaller fragments (< 1 kb), a substantial population of large fragments is still retained in Pu-Seq libraries made from ssDNA fragments that were not size selected. Large fragments in Illumina libraries do not cluster and sequence efficiently, which could reduce the quality of the data obtained from the library. It was noted that data from libraries that were made using size selected ssDNA showed the lowest background and were the most reproducible (Keszthelyi, Daigaku et al. 2015).

Seeing as the ligation of Illumina adaptors onto ssDNA is inefficient, the ssDNA fragments generated by the alkaline treatment are converted to dsDNA. The second (complementary) strands are synthesised from an equimolar pool of deoxyribonucleotide triphosphate (dNTPs), where deoxythymidine triphosphate (dTTP) was replaced with deoxyuridine triphosphate (dUTP), marking the newly synthesised strand with dUTPs. During the Illumina library preparation, after adaptor ligation step, the fragments are treated with Uracil-Specific Excision Reagent (USER) enzyme, which degrades the complementary strand that contained dUTPs. This ensured that only the originally isolated ssDNA fragments were amplified in the final library.

The polymerase δ and polymerase ε Pu-Seq libraries are then sequenced to a depth of ~10 million reads per library and the data are pooled and analysed together. The reads are aligned to a reference genome and binned into 300 bp windows. The 5' end of the reads in each bin are counted and normalised to the total number of reads. Assuming that each position can only be replicated by either polymerase δ or ε, the ratios of polymerase usage are calculated.

A more detailed protocol of the library preparation steps and subsequent data analysis is described in 2.2.4.

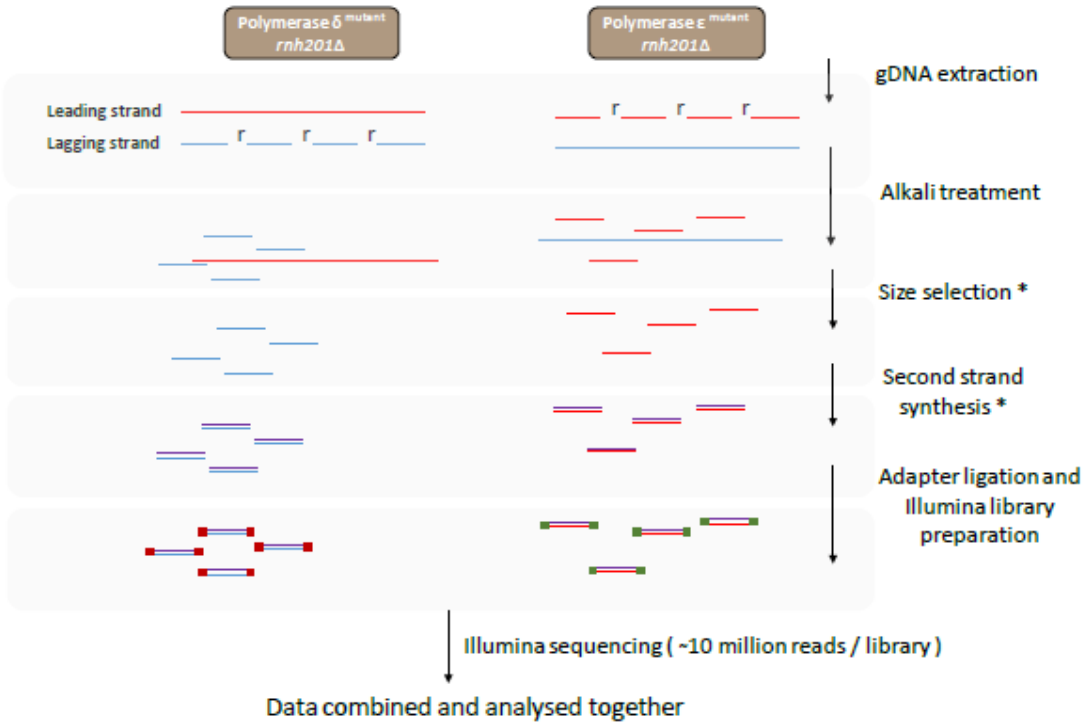


Fig. 3.2 – **Steps involved in the generation of Pu-Seq libraries**

The Pu-Seq polymerase δ (*cdc6-L591G rnh201*Δ) and polymerase ε (*cdc20-M603F rnh201*Δ) strains are each grown to mid-log phase and cells are collected for gDNA extraction. Cells with a mutation in the steric gate of polymerase δ, incorporate a higher than average amount of rNTPs in the lagging strand. The same is true for the leading strand in the polymerase ε mutant. gDNA is extracted from both cultures, treated with alkali to fragment at positions of rNTP incorporation and the ssDNA is size selected by excision from agarose gels. A complementary strand is synthesised and the fragments are used for Illumina library preparation. The libraries are barcoded, pooled and sequenced to a depth of ~ 10 million reads / library. Data from the polymerase δ and ε libraries are combined and analysed together.

Steps which have been optimised as part of work for this thesis, are marked with *.

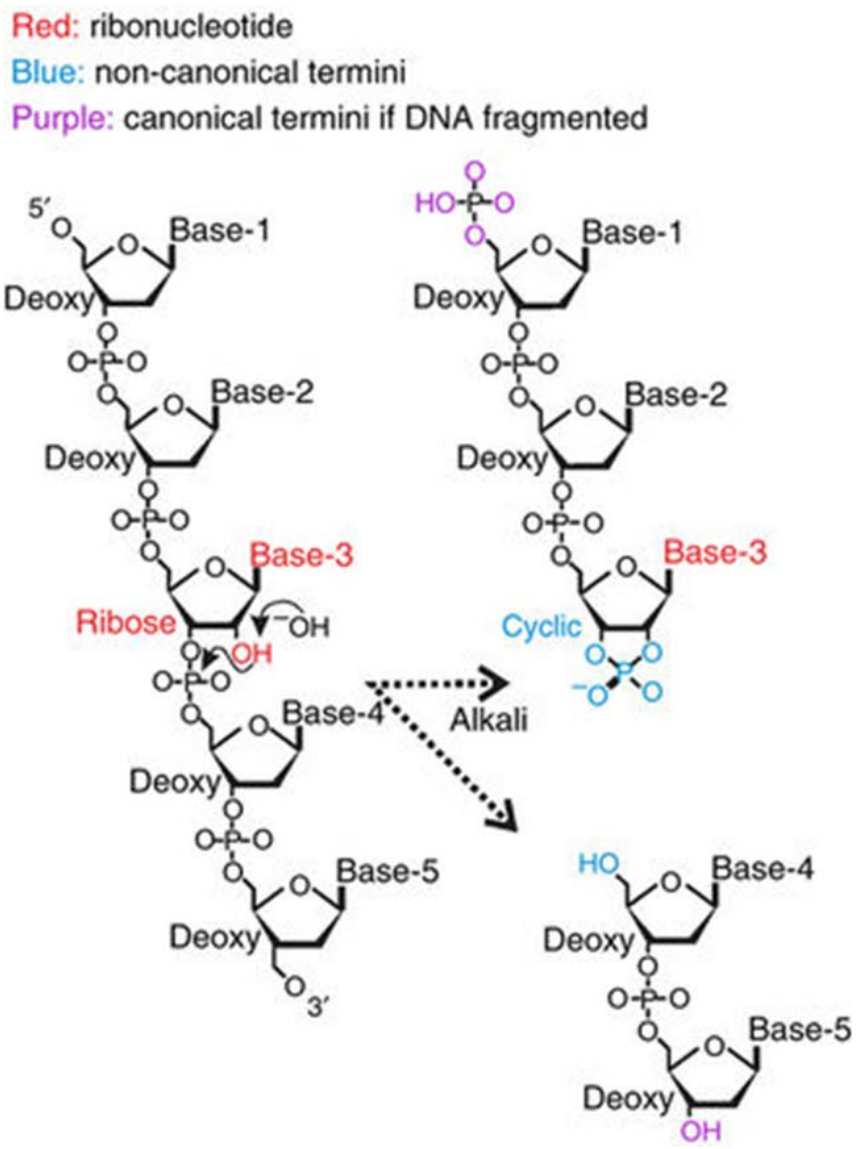


Fig. 3.3 – **Cleavage of a ribonucleotide incorporated in DNA by alkali**

The ribose sugar of a ribonucleotide has an additional 2'OH group (shown in red), compared to the deoxyribose sugars found in deoxyribonucleotides. A hydroxyl attack on the 2'OH group generates a cyclic phosphate (shown in blue) and leads to strand cleavage.

Adapted from (Keszthelyi, Daigaku et al. 2015).

3.1.3 Chapter aims

In order to analyse polymerase usage and bias at repetitive regions more accurately, we attempted to increase the coverage of repetitive regions. First, we optimised the steps leading up to the library preparation to generate higher quality libraries. We compared the activity of a number of polymerases at the second strand synthesis step, to improve its reproducibility. To analyse the impact of changing the polymerase at the second strand synthesis step on the quality of the data generated, we constructed Illumina libraries using the optimised protocol and compared the generated data to that from previous Pu-Seq experiments.

Next, we attempted a number of modifications to the original alignment algorithm, to directly improve the alignment of short reads across repetitive areas of the genome.

3.2 Results

3.2.1 Optimization of the second strand synthesis

Originally, the second strand synthesis was carried out using Invitrogen DNA Polymerase I, Large (Klenow) Fragment (referred to from now on as 'Klenow'), a polymerase with 3' to 5' proofreading activity. The reaction, however, was not very reproducible. The range of sizes of the dsDNA fragments produced and the total amount of dsDNA synthesized were very variable.

We set out to optimise this step by testing the activity of a number of other polymerases on both size selected and non size selected ssDNA. All of the polymerases tested had 3' to 5' proofreading activity and were active between 30°C and 37°C. The latter was necessary due to the low melting temperature of the 8N random primers used in the reaction.

3.2.1.1 Non size selected ssDNA

100 ng of non-size selected ssDNA was incubated with the polymerases for 40 minutes. The resultant dsDNA fragments were cleaned with Agencourt AMPure XP beads (Beckman Coulter) and analysed using a high sensitivity DNA chip on the Agilent 2100 Bioanalyser (Fig. 3.4 – polymerases indicated on the figure).

Two representative traces of the dsDNA fragments generated by the Klenow reaction are shown in Fig. 3.4. In one (Fig. 3.4 D), a moderate amount of dsDNA

fragments (~300 pg/µL) was distributed over a broad size range, ranging from 200 bp to over 1 kb. In the second (Fig. 3.4 E), there were no detectable dsDNA peaks. The reaction using NEB DNA Polymerase I, Large (Klenow) Fragment (referred to as Klenow (NEB)) generated much smaller fragments (peaking at ~200 bp), with a narrower distribution. Reactions using T4 and φ29 resulted in much higher concentrations of dsDNA fragments - 2 ng/µL and 1.4 ng/µL, respectively. Although the average sizes of the fragment generated were much larger (peaking at 3 kb and 2.5 kb, for T4 and φ29, respectively), the distribution of fragment sizes was much tighter. The reaction with T7 did not produce any detectable dsDNA fragments.

We wanted to determine whether the T4 and φ29 reactions could be modified to produce a narrow range of smaller fragments. To decrease the average fragment sizes, we repeated the incubation, using shorter polymerisation times (Fig. 3.5 – conditions indicated on the figure).

Decreasing the polymerisation time did not result in a greater population of smaller fragment sizes. Instead, all reactions generated lower concentrations of large fragments, with the concentrations of the dsDNA peaks decreasing with time. This suggests that both T4 and φ29 are most efficient at synthesising larger DNA fragments.

The data suggest that the second strand synthesis reaction using Klenow (NEB) produces dsDNA fragments with the most optimal distribution of sizes for Illumina library preparation. Replacing Klenow with Klenow (NEB) could, therefore, reduce the background in data generated from libraries made from non size selected ssDNA.

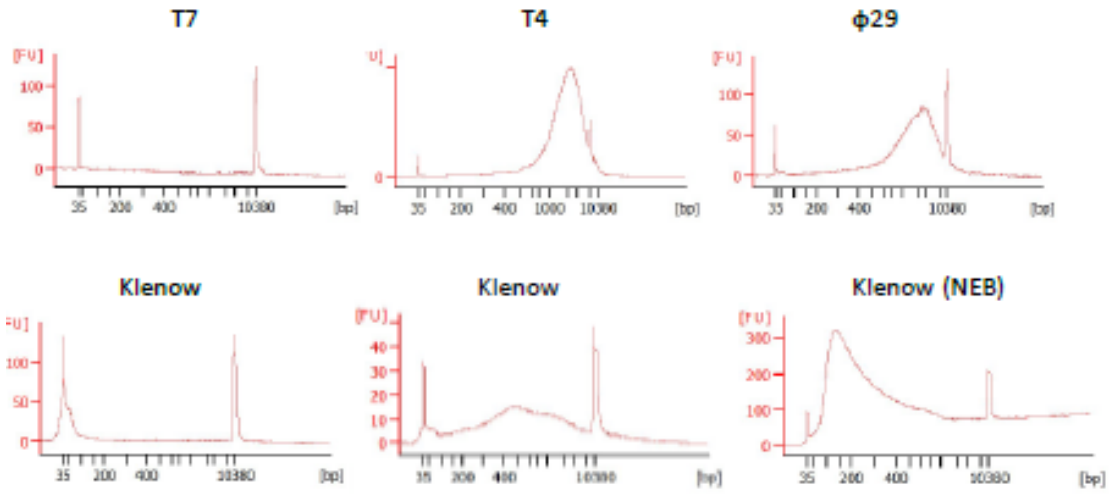


Fig. 3.4 – **Double stranded DNA fragments made from non size selected ssDNA using various polymerases**

100 ng of ssDNA was incubated for 40 minutes with 0.3 mg/mL (final concentration) 8N random primers, dNTP mix (where dTTP was substituted with dUTP – 2 mM each final concentration) and the polymerases indicated on the figure in appropriate buffers, at their respective optimum temperature (30°C for ϕ29 and 37°C for all other polymerases). The resultant dsDNA fragments were cleaned using Agencourt AMPure XP beads (1.8X) and 1 µl was loaded onto an Agilent High Sensitivity chip.

Lower and upper markers peak at 35 bp and 10380 bp, respectively. The peak in between the markers represents the size and amount of dsDNA fragments polymerised in each reaction.

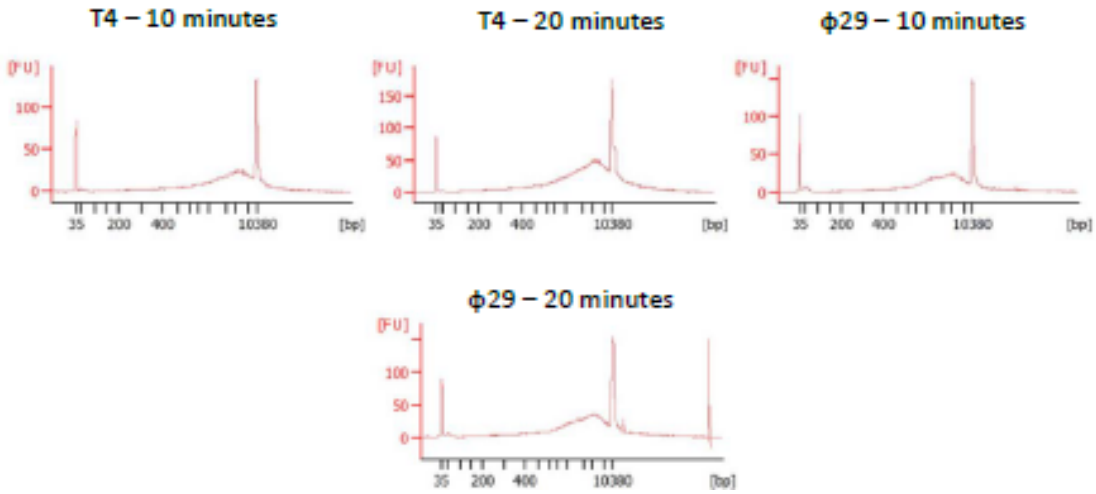


Fig. 3.5 – **Double stranded DNA fragments made from non size selected ssDNA using T4 and φ29 polymerases at altered conditions**

100 ng of ssDNA was incubated for the times indicated with 0.3 mg/mL (final concentration) 8N random primers, dNTP mix (where dTTP was substituted with dUTP – 2 mM each final concentration) and the polymerases indicated on the figure in appropriate buffers, at their respective optimum temperature (30°C for ϕ29 and 37°C for T4). The resultant dsDNA fragments were cleaned using Agencourt AMPure XP beads (1.8X) and 1 µl was loaded onto an Agilent High Sensitivity chip.

Lower and upper markers peak at 35 bp and 10380 bp, respectively. The peak in between the markers represents the size and amount of dsDNA fragments polymerised in each reaction.

3.2.1.2 Size selected ssDNA

100 ng of size selected DNA was incubated with the polymerases for 20 minutes and the resultant dsDNA was cleaned and analysed as described in 3.2.1.1 (Fig. 3.6 - polymerases indicated on the figure). The reaction with Klenow was not very efficient, producing a broad dsDNA peak of 100 pg/µL. The dsDNA fragments produced, however, were all within the range that could be used for library construction (300 bp - 500 bp). Again, the reactions using T4 and φ29 resulted in the highest concentration of dsDNA fragments (1 ng/µL and 500 pg/µL, respectively) distributed in a narrow range (150 bp - 500 bp). The reaction using T7 resulted in a small amount (50 pg/µL) of narrowly distributed dsDNA fragments (peak at ~350 bp). Klenow (NEB) did not generate dsDNA fragments whose sizes were normally distributed - two peaks were seen at ~ 125 bp and 300 bp and the fragment sizes extended to over 700 bp.

The second strand synthesis reaction using T4 and size selected ssDNA resulted in the highest concentration of dsDNA fragments in appropriate size ranges for Illumina library construction. Given the better efficiency and reproducibility compared to those seen when using Klenow, T4 was subsequently used to replace Klenow in the second strand synthesis reaction of size selected ssDNA.

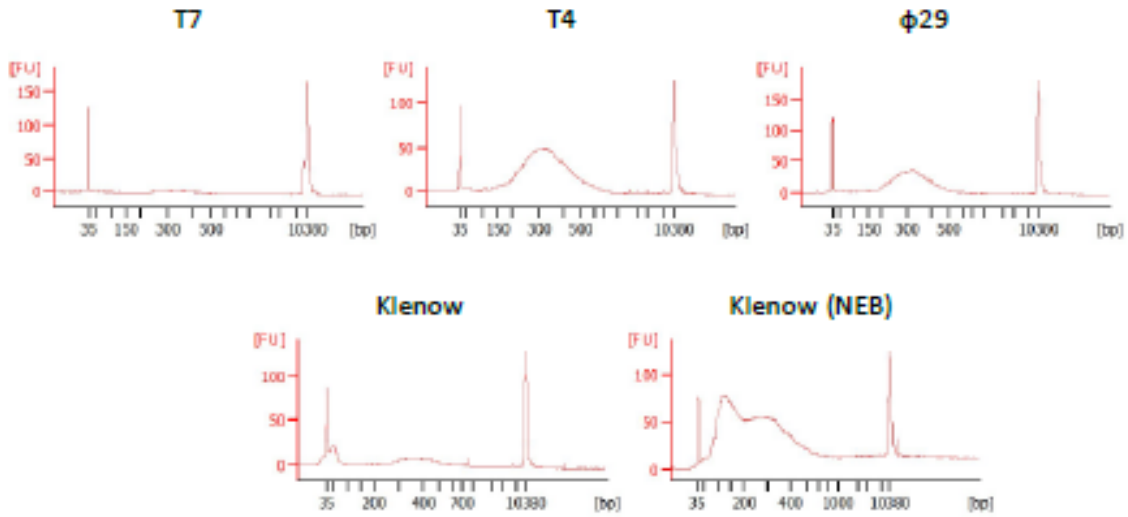


Fig. 3.6 – **Double stranded DNA fragments made from size selected ssDNA using various polymerases**

100 ng of ssDNA was incubated for 20 minutes with 0.3 mg/mL (final concentration) 8N random primers, dNTP mix (where dTTP was substituted with dUTP – 2 mM each final concentration) and the polymerases indicated on the figure in appropriate buffers, at their respective optimum temperature (30°C for ϕ29 and 37°C for all other polymerases). The resultant dsDNA fragments were cleaned using Agencourt AMPure XP beads (1.8X) and 1 µl was loaded onto an Agilent High Sensitivity chip.

Lower and upper markers peak at 35 bp and 10380 bp, respectively. The peak in between the markers represents the size and amount of dsDNA fragments polymerised in each reaction.

3.2.2 Constructing and sequencing new wild type T4 Pu-Seq libraries

To determine whether the new protocol minimized the noise in the data, we constructed Pu-Seq libraries using the optimized protocol with T4 and size selected ssDNA. The dsDNA generated were subsequently used to generate Illumina libraries (Fig. 3.7).

The libraries were amplified using unique index primers and mixed for multiplex paired-end sequencing to generate ~10 million reads per library. The paired-end reads were aligned to the SP2 reference genome using Bowtie2, which was set to perform end-to-end alignment in mixed mode. The end-to-end alignment option ensured that all of the read characteristics were used during the alignment, i.e., it did not allow for the clipping of reads to maximise the alignment score. Mixed mode allowed for the single-end mapping of reads when a paired-end alignment was not possible. Additionally, to improve the average read quality, the 101 bp long paired-end reads were trimmed 30 bp at the 3' end and 1 bp from the 5'end. The combination of these Bowtie2 settings will be referred to as the "standard alignment".

The generated data were compared to those from previously analysed wild type Pu-Seq libraries that were made using Klenow at the second strand synthesis step.

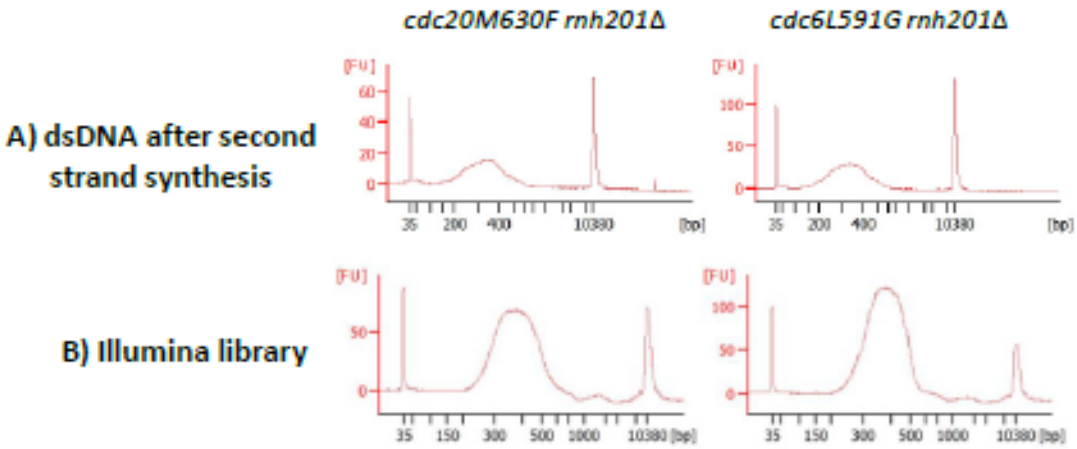


Fig. 3.7 – **Double stranded DNA fragments generated by the optimised second strand synthesis reaction and the resultant Illumina libraries**

A) Second strand synthesis. 100 ng of ssDNA was incubated for 20 minutes with 0.3 mg/mL (final concentration) 8N random primers, dNTP mix (where dTTP was substituted with dUTP – 2 mM each final concentration) and T4 DNA polymerases in NEB 2.1 buffer at 37°C. The resultant dsDNA fragments were cleaned using Agencourt AMPure XP beads (1.8X) and 1 µl was loaded onto an Agilent High Sensitivity chip. Lower and upper markers peak at 35 bp and 10380 bp, respectively.

B) Illumina libraries. 50 µl of the dsDNA fragments were used to construct an Illumina library, using the NEBNext Ultra library prep kit, following manufacturer's instructions. The libraries were diluted 1 in 5 in ultrapure water and 1 µl was loaded onto an Agilent High Sensitivity chip. Lower and upper markers peak at 35 bp and 10380 bp, respectively.

3.2.3 Comparison of T4 and Klenow datasets

The alignment rate and coverage by the reads generated from the Pu-Seq libraries made using Klenow and T4 fragments are compared in Table 3.1. The number of reads from the Klenow libraries varied between 16 and 25 million reads. This difference could have been caused by an error during the quantification and pooling of the libraries. The low read number of the Klenow polymerase ε library, however, correlates with a low total alignment rate. This suggests that the variable read number could have been caused by the overall low quality of the library. Reads from the T4 libraries, on the other hand, both generated approximately 12 million reads each. They also had very similar alignment rates, which were both greater than either of the alignment rates of the reads from the Klenow libraries.

Based on the total numbers of aligned reads, the coverage of the 12.57 Mb haploid genome was calculated using the Lander/Waterman equation (Lander and Waterman 1988), as recommended by Illumina:

Coverage = (read length (bp) * number of reads) / size of haploid genome (bp)

Given the varying number of reads and alignment rates, the coverage by the reads from the Klenow libraries ranged from between ~58 and ~100 (Table 3.1). The coverage by the reads from the T4 libraries was ~70 for both libraries. This was higher than that for the Klenow polymerase ε library, despite the Klenow library generating over 4 million more reads.

Table 3.1 – **Number of reads obtained and mapped for the libraries made using Klenow and T4 fragments**

The overall alignment rate represents the number of mates that aligned in pairs (concordantly and discordantly) and single reads that aligned in single end mode.

Polymerase used during second strand synthesis	Replicative polymerase mutated in the strain	Total number of reads	Overall alignment rate	Number of reads mapped to the reference genome	Coverage
Klenow (Invitrogen)	Polymerase δ	25,655,294	74.48%	19,108,062	121.6
	Polymerase ε	16,349,242	55.84%	9,129,416	58.1
T4	Polymerase δ	12,801,495	90.14%	11,539,267	73.4
	Polymerase ε	12,482,564	87.07%	10,868,568	69.2

The aligned reads were processed using the same pipeline of Perl and R scripts (described in detail in 2.2.4.3). The global polymerase usage is calculated as a ratio between the normalised counts contributed by each polymerase and the total number of normalised counts contributed by both polymerases, at each bin. Given that the traces of polymerase usage are based on this ratio, data from both the T4 and Klenow experiments generated the same global patterns of polymerase usage (polymerase usage across Chromosome 3 is shown in Fig. 3.8). The values of the ratios generated by the T4 data, however, were higher. For example, the mean usage of polymerase δ on the forward strand (on a scale from 0 to 1) ranged from 0.36 to 0.65 for the Klenow data and 0.31 to 0.71 for the T4 data. This was likely caused by the reduced noise in the T4 data.

The increased reproducibility of the reactions upstream of the library preparation, as well as the increased coverage did not, however, remove all of the variability between datasets. The progression of leftward moving forks generated by three independent wild type T4 Pu-Seq library sets was compared using a Kolmogorov-Smirnov test (KS). The KS-test measures the cumulative difference between two continuous variables (expressed as a D statistic) and the probability that they are statistically different. Comparing the data from two of the three datasets showed a very low cumulative difference and a high p-value (D = 0.0041042, p-value = 0.872), suggesting that they are statistically similar. Comparison with the third dataset, however, resulted in a much higher cumulative difference and a lower corresponding p-value (D = 0.011239, p-value = 0.01005). This variability is one of the current limitations of the analysis and has to be taken into account when small differences between fork progression and polymerase bias are being considered. For future work and more accurate comparisons, it is important to quantify this variability and express a range of possible values for all of the parameters in a wild type background. This would allow to confidently identify differences when comparing data between wild type and mutant backgrounds.

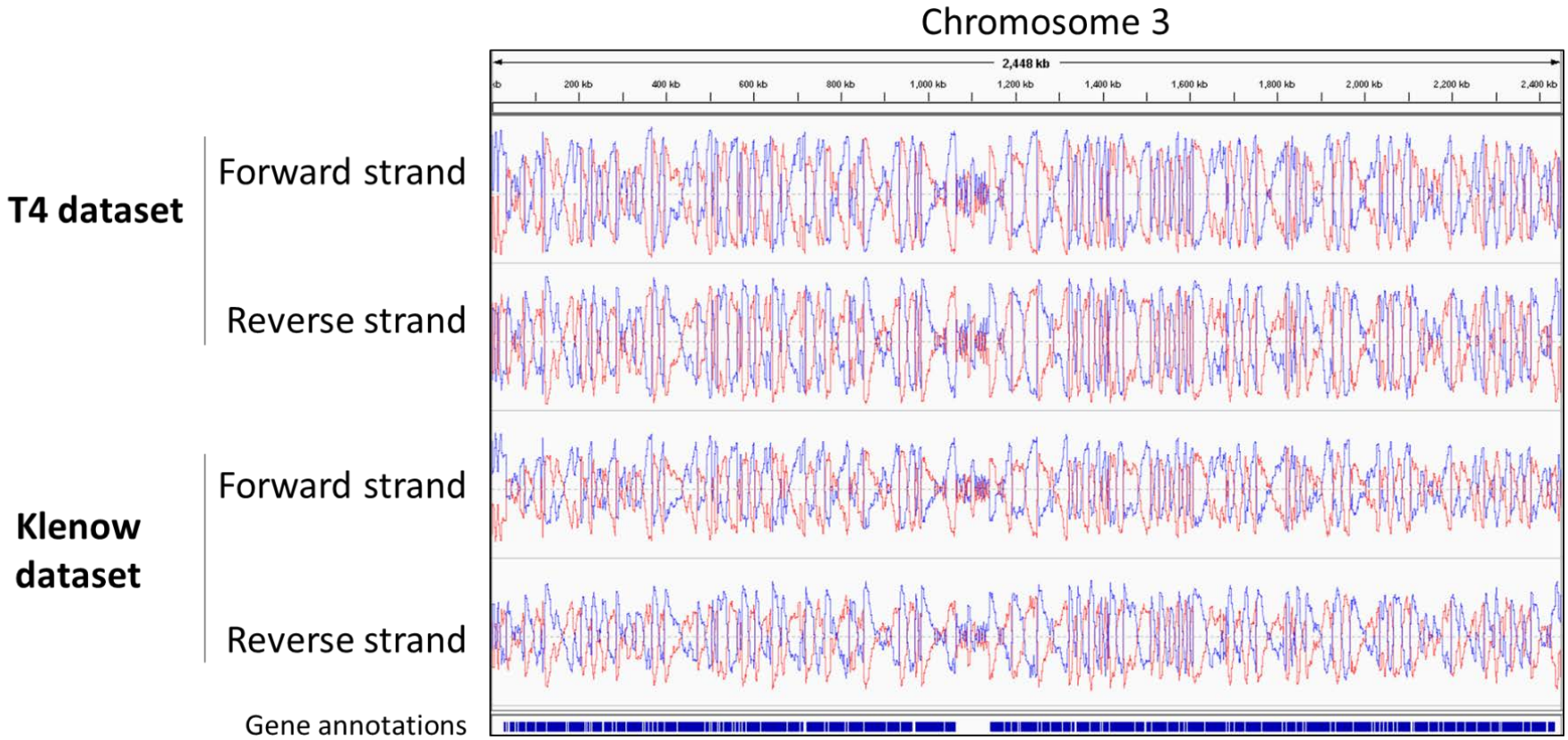

Chromosome 3
2,448 kb
200 kb
400 kb
600 kb
800 kb
1,000 kb
1,200 kb
1,400 kb
1,600 kb
1,800 kb
2,000 kb
2,200 kb
2,400 kb
T4 dataset
Forward strand
Reverse strand
Klenow dataset
Forward strand
Reverse strand
Gene annotations

Fig. 3.8 – **Polymerase usage across Chromosome 3, calculated from data generated by libraries that were constructed using polymerases T4 and Klenow at the complementarsy strand synthesis stage**

Reads generated by the libraries indicated on the figure were trimmed (30 bp at the 3' end and 1 bp at the 5' end) and aligned to the SP2 reference genome. The reads were binned into 300 bp windows and the number of reads in each bin was normalised to the total number of reads. Polymerase usage was calculated by taking the ratio of the normalised counts contributed by each polymerase to the total number of normalised counts contributed by both polymerases. The usage of polymerases δ and ε (shown in blue and red, respectively) on the forward and reverse strands were visualised using the Integrative Genome Viewer (IGV) genome browser.

3.2.4 Optimising the alignment algorithm

In addition to the variability between datasets, the poor alignment of short reads to repetitive regions poses another problem to an accurate description of polymerase usage at repetitive regions. Due to the heuristic nature of the Bowtie2 algorithm, if the generated alignment score for a read is the same for a number of loci, the read will be mapped to one of those loci at random. This will be reflected in a low mapping quality score, which represents the uniqueness of the alignment.

Genome wide mapping quality of the standard alignment was assessed using Qualimap 2 (Okonechnikov, Conesa et al. 2016). All three of the centromeric regions had considerably lower mapping quality score compared to the rest of the genome (Fig. 3.9). Centromere 1 had the highest mapping quality of all three centromeres, suggesting that the *imr* and *cnt* regions of Centromere 1 may have been aligned accurately and the lower quality was caused by the multiple alignment possibilities of the reads originating from the *otr* repeats. To improve the overall alignment to the centromere, we attempted to increase the mapping quality around the *otr* repeats.

One way to reduce the possibility of multiple mapping is to increase the read length. In the standard alignment, reads were trimmed to 80 bp, as their quality deteriorated towards the 3' end. In order to maintain the high quality of reads while allowing for reads longer that 80 bp, we adjusted the trimming to each read individually. This was done using the trimming algorithm Trimmomatic (Bolger, Lohse et al. 2014). Trimmomatic scans each read, trimming bases from the 3' end until the average read quality rises above the set quality value (Q). The same set of paired-end reads (FASTQ format) were trimmed to a range of Q values and a minimum length of 36 bp. The resultant reads were analysed using FastQC. The average quality of the reads, especially towards the 3' end, increased with increasing Q value (Fig. 3.10 A). The number of reads remaining after trimming (as a percentage of the total number of reads) is shown in Fig. 3.10 B.

Next, we aligned the reads, trimmed to different Q values using Bowtie2. To further reduce the possibility of multiple mapping we also modified the alignment algorithm. Multi-seed algorithms, such as Bowtie2, generate substrings (or seeds) of a defined length in each read, which they align to set criteria. The alignment is

then extended to the full length of the read. Increasing seed length, reducing the number of mismatches allowed per seed, as well as increasing the number of attempts to extend a seed (or re-seed a read) can all be employed to increase alignment quality across repetitive regions. The Bowtie2 "--very-sensitive" parameter, collectively changes all of these running options.

The trimmed reads were aligned to the SP2 reference genome using Bowtie2 --very-sensitive end-to-end alignment in mixed mode. The SAM files generated by the Bowtie2 alignment were converted into BAM format, sorted and viewed using the genome browser IGV (Integrative Genomics Viewer). The coverage for the region assigned as Centromere 1 in SP2 (Chr1:3,753,687 - 3,789,421) is shown in Fig. 3.11. The coverage was highest at the outer regions of the centromere, suggesting that the reads from the *dg* repeats from the other centromeres may have been mapped to Centromere 1. The coverage of the *imr* and *cnt* regions of Centromere 1 was low for of the individually trimmed reads aligned using the --very-sensitive option. Overall, there was no visible difference between the coverage compared to the original algorithm (80 bp reads aligned using standard end-to-end alignment), suggesting that the modification of the read trimming and the alignment algorithm does not significantly increase the quality of the data at Centromere 1.

To determine whether the new algorithm resulted in a change in the genome wide mapping quality, we compared the alignment of the trimmed reads that were aligned using the --very-sensitive option to the standard algorithm using the multi-sample BAM QC analysis in Qualimap 2. The genome wide mapping quality of the trimmed reads was lower, but still comparable to the standard alignment (Fig. 3.12).

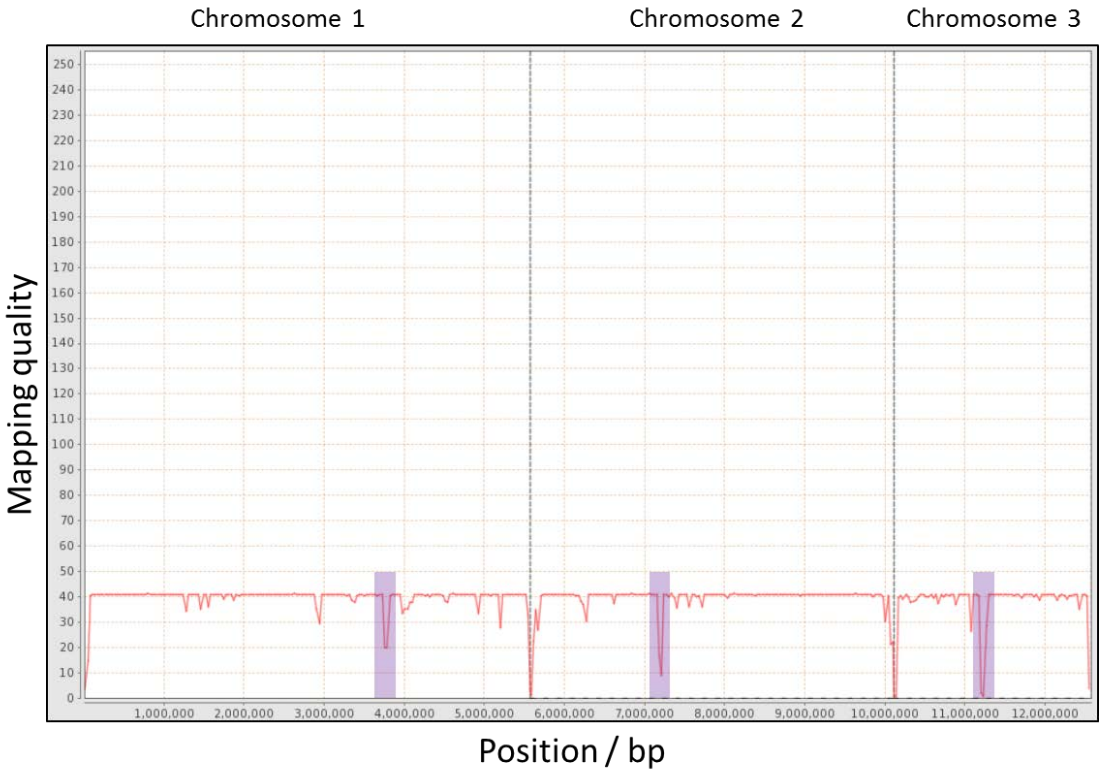
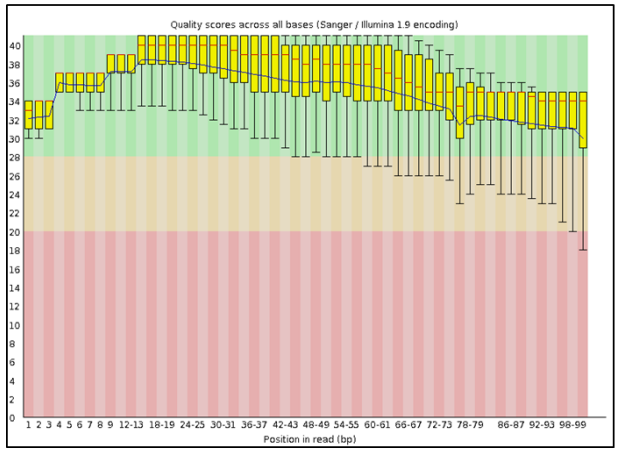


Fig. 3.9 - **Genome wide mapping quality of reads aligned using Bowtie2, generated by Qualimap**

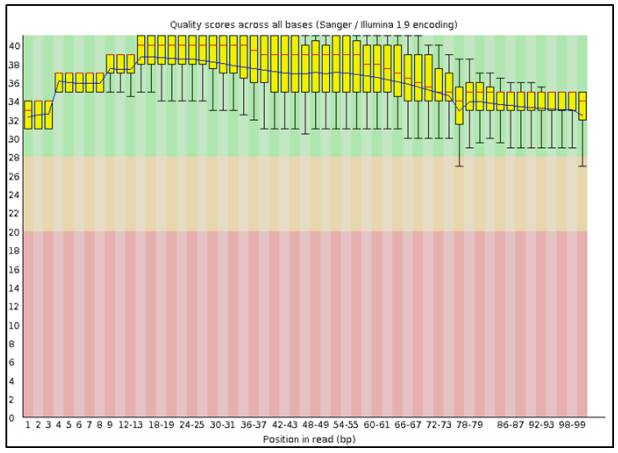
101 bp long paired-end reads were trimmed 30 bp at the 3' end and 1 bp at the 5' and aligned using Bowtie2 end-to-end alignment in mixed mode. The generated SAM file was analysed using Qualimap 2. Dips in mapping quality were seen around centromeres and telomeres (centromeres are marked in purple).

A) FastQC analysis of reads

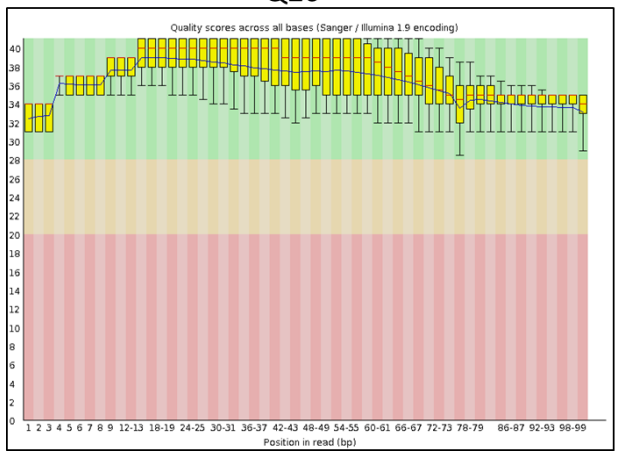
No trimming



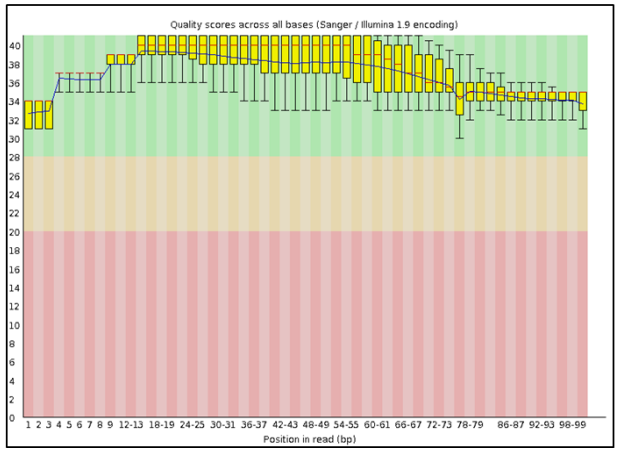
Q15



Q20



Q25



B) Percentage of reads after trimming

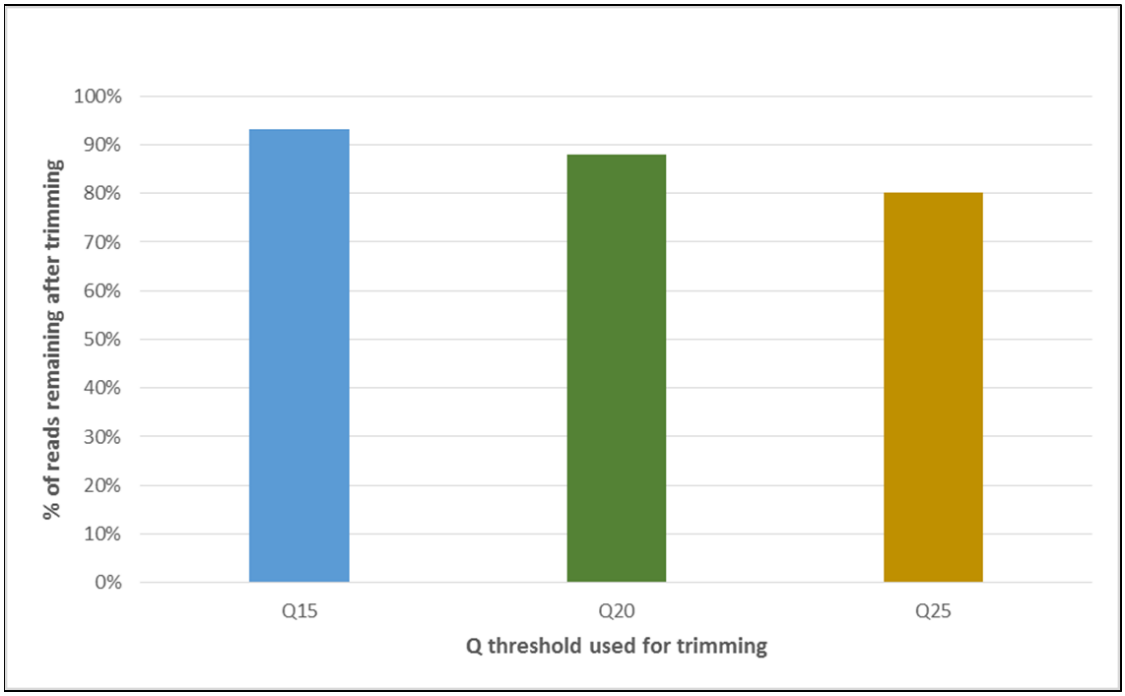


Fig. 3.10 - **Sequence quality and percentage of reads surviving after trimming to a range of Q values**

One set of paired end reads was trimmed, using paired end Trimmomatic. A sliding window of 4 bp was set to scan and trim each read until the average quality rose above the Q values indicated on the figures. A minimum read length of 36 bp was set. Data for one of the paired end mates is shown.

A) FastQC analysis. The resultant reads were analysed using FastQC and the per base sequence quality is shown.

B) Percentage of the number of reads surviving each trimming to the total number of reads

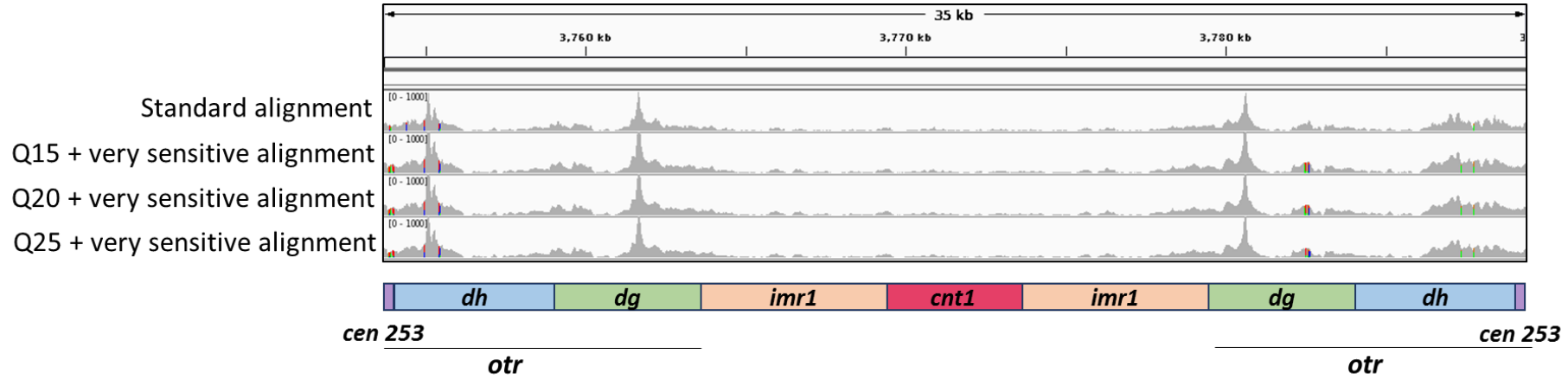


Fig. 3.11- **Coverage of Centromere 1 using the modified and standard alignment algorithms**

One set of paired end reads was trimmed, using paired end Trimmomatic. A sliding window of 4 bp was set to scan and trim each read until the average quality rose above the Q values indicated on the figures. A minimum read length of 36 bp was set. The trimmed reads were aligned to the SP2 reference genome using Bowtie2 --very-sensitive end-to-end alignment in mixed mode. The same reads were also aligned using the standard alignment algorithm.

The generated SAM files were converted to BAM format, sorted, indexed and viewed on the Integrative Genome Viewer (IGV) genome browser. Coverage of the region assigned as Centromere 1 in the SP2 reference genome (Chr1: 3,753,687 - 3,789,421) is shown. A schematic of the centromere organisation is shown below to indicate the coverage over the central and outer regions.

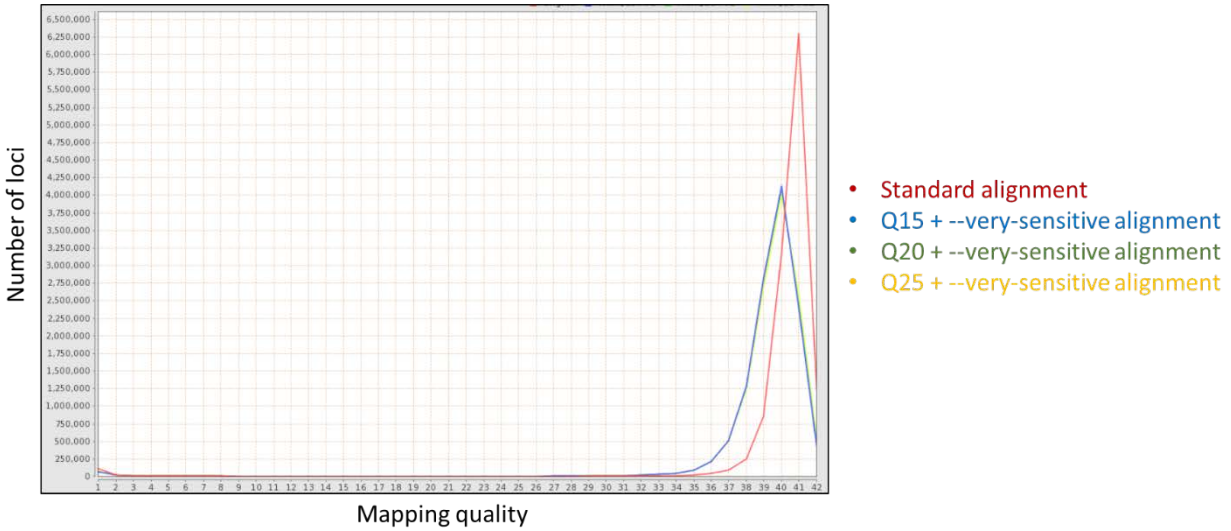


Fig. 3.12 - **Genome wide mapping quality using the modified and standard alignment algorithms**

One set of paired end reads was trimmed, using paired end Trimmomatic. A sliding window of 4 bp was set to scan and trim each read until the average quality rose above the Q values indicated on the figure. A minimum read length of 36 bp was set.

The trimmed reads were aligned to the SP2 reference genome using Bowtie2 --very-sensitive end-to-end alignment in mixed mode. The same reads were also aligned using the standard alignment. The generated SAM files were converted to BAM format and compared using the multi-sample BAM QC analysis in Qualimap 2. The mapping quality histogram is shown. The mapping quality of the reads trimmed using Trimmomatic were almost identical and may be difficult to discern on the figure.

3.3 Discussion

3.3.1 Optimisation of the library preparation

The second strand synthesis reaction using non size selected ssDNA was improved by using Klenow (NEB) (Fig. 3.4). The reaction generated small fragments that peaked at 134 bp and were all in a size range that could be used for Illumina library preparation. The reactions using T4 and φ29 produced longer fragments (over 1.5 kb) and considerably more dsDNA than the Klenow reaction (Fig. 3.4). Deceasing the polymerisation time resulted in lower amounts of long dsDNA fragments, instead of shorter fragments (Fig. 3.5), suggesting that T4 and φ29 are more efficient than Klenow in the polymerisation of the complementary strand of long ssDNA templates. The reactions, however, did not generate fragments in the size range (150 - 500 bp) that is required for Illumina library preparation.

In the presence of only short ssDNA templates (size selected ssDNA), all of the tested polymerases generated varying amounts of small dsDNA fragments (Fig. 3.6). The reaction with T4 synthesised the complementary strand with the highest efficiency. Overall, despite the optimisation with Klenow (NEB), size selected ssDNA is a better substrate for second strand synthesis and has been reported to reduce background (Keszthelyi, Daigaku et al. 2015). Size selection was, therefore, included as a non-optional step in the preparation of Pu-Seq libraries to study repetitive regions.

The dsDNA fragments synthesised by T4 from size selected ssDNA were subsequently used to prepare wild type Pu-Seq polymerase δ and ε libraries (Fig. 3.7). The libraries generated over 12 million reads each, most of which aligned well to the SP2 reference genome and resulted in a coverage of ~70 (Table 3.1).
The wild type polymerase δ and ε Pu-Seq libraries made using fragments synthesised by Klenow generated over 25 and 16 million reads, respectively (Table 3.1). This difference could be explained by an error during the quantification and pooling of the libraries before the sequencing. That would, however, not explain the very different alignment rates (Table 3.1). It is possible that the variability in the number and the sizes of the fragments generated by the Klenow reaction could have impacted the quality of the final library.

Given that the Pu-Seq polymerase ratios are calculated from normalised data, the polymerase usage data generated from the T4 libraries resulted in the same pattern of polymerase usage as that from Klenow data. The pattern, however, was less noisy in the T4 dataset (Fig. 3.8). Taken together, the data show that generating the complementary strands using size selected ssDNA and T4 has a considerable impact on the quality of the libraries generated and the data they produce. We have, therefore, adapted the protocol to include these modifications.

3.3.2 Optimisation of the alignment algorithm

The individual trimming of reads to set Q values by Trimmomatic increased the average read quality, compared to untrimmed reads (Fig. 3.10 A). The alignment of these reads (using the --very-sensitive option in Bowtie2), however, did not result in an increased coverage of Centromere 1 (Fig. 3.11). The genome wide mapping quality increased with increasing Q value, but in every case was lower than that for the standard alignment (Fig. 3.12). Based on these data, the modification of the read trimming and the alignment algorithm does not strikingly improve the coverage and mapping quality around Centromere 1.

Future analyses could be improved by using longer reads. Third generation sequencing platforms, such as Oxford Nanopore Technologies allow the generation of reads up to several hundred kilo bases (Lu, Giordano et al. 2016). This would allow an accurate *de novo* assembly of all of the centromeric and telomeric regions. The long reads would also align better with much higher mapping quality to the repetitive regions, increasing coverage and mapping quality.

Given the current restrictions to a sub optimal reference genome and short reads, we decided not to change the mapping algorithm. The poor coverage of Centromere 1 was noted as a caveat to any analyses carried out around that region.

Chapter 4

4.1 Introduction

Similarly to other eukaryotes, *S. pombe* heterochromatin can be subdivided into constitutive and facultative heterochromatin. The pathways specific to the formation of each type of heterochromatin in *S. pombe* are briefly described below.

4.1.1 Constitutive heterochromatin in *S. pombe*

In *S. pombe*, constitutive heterochromatin is assembled at centromeres, telomeres, rDNA and the *mat* locus (Cam, Sugiyama et al. 2005). The formation of constitutive heterochromatin is driven by an RNAi dependent mechanism, described in 1.3.2 (Fig. 1.3). Briefly, long RNAs transcribed from *dg* and *dh* repeats are processed into short siRNAs and associate with the RITS complex. The siRNAs then guide the RITS complex back to the loci from which they were transcribed, starting the cascade of Clr4 dependent histone H3K9 methylation, which in turn recruits Swi6 to the heterochromatin regions (Grewal and Elgin 2007).

The deletion of the Chp1, the H3K9me binding chromodomain component of RITS (Verdel, Jia et al. 2004), decreases histone methylation at heterochromatin regions to varying degrees in each of the domains (Sadaie, Iida et al. 2004). This suggests that the formation of constitutive heterochromatin at each of these regions has additional RNAi independent heterochromatin targeting mechanisms.

DNA binding proteins have been shown to be involved in nucleating heterochromatin at centromeres, telomeres and the *mat* locus. The binding of ATF/CREB family proteins Atf1 and Pcr1 at the *mat* locus creates a nucleation site for heterochromatin by interacting directly with Clr4 and Swi6 (Jia, Noma et al. 2004). Similarly, Taz1, can nucleate heterochromatin formation at subtelomeres independently of the RNAi pathway (Kanoh, Sadaie et al. 2005). It has also been shown that the deletion of Abp1, Cbh1, and Cbh2 (*h*CENP-B), diminishes centromere heterochromatin (Nakagawa, Lee et al. 2002).

4.1.2 Facultative heterochromatin in *S. pombe*

Facultative heterochromatin in *S. pombe* is found in 21 discreet blocks along the genome. These blocks (or islands) of increased H3K9 methylation have been subdivided into two classes. The first class of the facultative heterochromatin islands are associated with meiotic genes and silence their expression during

vegetative growth (Zofall, Yamanaka et al. 2012). The proteins and elements involved in the maintenance of facultative heterochromatin at these sites are also involved in silencing the expression of the meiotic genes at the transcriptional level. RNA-binding protein Mmi1 binds to mRNA transcripts that contain a DSR (determinant of selective removal) element, a *cis*-acting region associated with meiotic genes, marking them for degradation by a nuclear exosome (Harigaya, Tanaka et al. 2006). Mmi1 also interacts with Erh1, which recruits the MTRC (Mtl1-Red1 core) complex required for the nucleation of heterochromatin (Sugiyama, Thillainadesan et al. 2016). The DSR region has also been reported to interact with another *cis*-acting element associated with the *mei4* gene (which codes for a meiotic transcription factor), further promoting H3K9me around the gene. (Tashiro, Asano et al. 2013). Although this effect was not seen at other meiotic genes, it highlights the importance of *cis*-acting elements in the formation of facultative heterochromatin.

The second class of heterochromatin islands are found at genes that are not associated with a DSR element. A number of these islands remain heterochromatic during nitrogen starvation, suggesting that they are not all required for meiotic gene regulation (Zofall, Yamanaka et al. 2012, Zofall, Smith et al. 2016). Neither the pathway of heterochromatin formation nor a non-meiotic role for these 6 non-DSR islands has been fully described. It is known that the formation of H3K9me is controlled by the binding of Taz1 (Zofall, Smith et al. 2016), which is reminiscent of its role in RNAi independent heterochromatin formation at telomeres (Kanoh, Sadaie et al. 2005). Taz1 has been shown to recruit Rif1 to these (Taz1 dependent) facultative heterochromatin islands (Zofall, Smith et al. 2016). The binding of Taz1 and Rif1 to these islands is necessary but not sufficient for the formation of heterochromatin, as they bind at other loci without this effect. Interestingly, 5 out of the 6 Taz1-dependent heterochromatin islands were discovered independently as Rif1 and Taz1 dependent *cis*-acting boundaries between euchromatin and heterochromatin (Toteva, Mason et al. 2017). This suggests that while they may have some overlapping role in meiotic gene repression with DSR islands, it is possible that Taz1 dependent heterochromatin islands may also have a separate role in the maintenance of gene expression and genomic stability.

Despite the differences, however, both classes of heterochromatin islands are affected by heterochromatin destabiliser Epe1. Epe1 is recruited to heterochromatin regions via Swi6 (Zofall and Grewal 2006), suggesting that the dynamic formation and destabilisation of heterochromatic marks are important for the formation of wild type levels of heterochromatin. Overexpression of Epe1 leads to an increase in euchromatin specific histone acetylation (Ayoub, Noma et al. 2003), while its loss increases histone methylation associated with heterochromatin (Zofall, Yamanaka et al. 2012). The increase in H3K9 methylation manifested in the spreading of the H3K9me3 marks around facultative islands, as well as the appearance of over 30 additional peaks of H3K9me (Zofall, Yamanaka et al. 2012). Interestingly, in *epe1*Δ all of the facultative heterochromatin islands persist longer in the absence nitrogen (Zofall, Yamanaka et al. 2012), suggesting that despite being established using different pathways, the heterochromatin formed is the same across all of the islands.

4.1.3 Chapter aims

As discussed in Chapter 3, heterochromatin has been suggested to impede fork progression (Rozenzhak, Mejia-Ramirez et al. 2010). This chapter will discuss the impact of heterochromatin found on replication dynamics measured using Pu-Seq.

We analysed the effects of both abolishing and increasing the levels of heterochromatin genome wide on origin firing and the progression of replication forks around Centromere 1. Using a bias towards polymerase δ usage as a marker of HR restarted replication, we attempted to identify whether wild type or modified levels of heterochromatin can act as RFBs in that region.

Next, we compared how affecting heterochromatin globally (using *swi6*Δ, *clr4*Δ and *epe1*Δ mutants) and only at a subset of facultative heterochromatin islands (using *rif1*Δ and *taz1*Δ mutants) impacted replication dynamics.

4.2 Results

4.2.1 Changes to replication dynamics in constitutive heterochromatin

To determine the impact of heterochromatin, we analysed replication dynamics in *swi6*Δ and *clr4*Δ backgrounds using Pu-Seq, assuming that all heterochromatin would be lost in these strains. We also analysed the effect of an increased heterochromatin load by analysing *epe1*Δ strains.

4.2.1.1 Replication dynamics in *swi6*Δ and *clr4*Δ

The number of origins mapped in a wild type background ranged between 999 and 1122, with an average of 1070 origins (data from 5 wild type Pu-Seq experiments). The number of origins mapped in the *swi6*Δ and *clr4*Δ strains was within this wild type range (*swi6*Δ – 1018 origins; *clr4*Δ – 1027 origins). It has been previously reported that in wild type *S. pombe*, the distribution of origin firing efficiencies is bimodal, with most origins firing inefficiently (with less than 40% efficiency) (Daigaku, Keszthelyi et al. 2015). We did not observe any notable changes to the distribution of the origin firing efficiencies in *swi6*Δ or *clr4*Δ (Fig. 4.1).

Despite the lack of an effect on global origin firing activity, we compared the progression of replication forks and origin usage around Centromere 1. The landscape of origin usage and the pattern of replication fork progression around the centromere were reproducible in a wild type background (Fig. 4.2 A).

Compared to wild type, more origins were mapped to the centromeric region in *swi6*Δ (Fig. 4.2 A) and most of these origins fired inefficiently. As a result of the effective decrease in origin firing efficiency within the centromere, fewer leftward moving replication forks originated from inside the centromere. A similar effect was seen on the right hand side of the centromere (changes in fork progression at the centromere are marked in orange on Fig. 4.2 A). Additionally, the decrease in the efficiency of firing of an origin at the edge of the right-hand portion of the centromere (marked by * on Fig. 4.2 A) correlated with an increase in the number of leftward moving forks originating outside the centromere in a region ~30 kb downstream of the centromere.

The changes in the origin landscape and pattern of replication fork progression were less pronounced in *clr4*Δ compared to *swi6*Δ. This difference in phenotype

was unexpected given the previous reports showing that the localisation of Swi6 to the outer regions of Centromere 1 is dependent on Clr4 (Ekwall, Nimmo et al. 1996, Partridge, Borgstrom et al. 2000).

Given the differences in the landscape of origin usage in the absence of heterochromatin (Fig. 4.2 A), another caveat had to be considered when analysing the polymerase bias (as a marker of HR restarted replication) around Centromere 1. As discussed in 1.2.2, recent data suggest that polymerase δ may contribute to leading strand replication around origins before polymerase ε replication is established (Daigaku, Keszthelyi et al. 2015, Yeeles, Janska et al. 2017, Garbacz, Lujan et al. 2018). In *S. pombe*, a bias of polymerase δ usage has been reported around efficiently firing origins of replication (Daigaku, Keszthelyi et al. 2015). In a population of cells, the replication of a region where replication forks are prone to stalling and collapse is likely to be rescued by a combination of HR restarted replication and dormant origin firing. Given that both of these can contribute to a bias of polymerase δ usage and Pu-Seq data describe the average usage of polymerases in a population, it would be impossible to differentiate between the causes of the bias.

The bias of polymerase δ usage was not reproducible between two wild type strains – Fig. 4.2 B, panels 3 and 4. Although *swi6*Δ and *clr4*Δ did show some differences, compared to either of the wild type data sets and to each other (Fig. 4.2 B, panels 1 and 2), it was impossible to attribute these differences to any one factor.

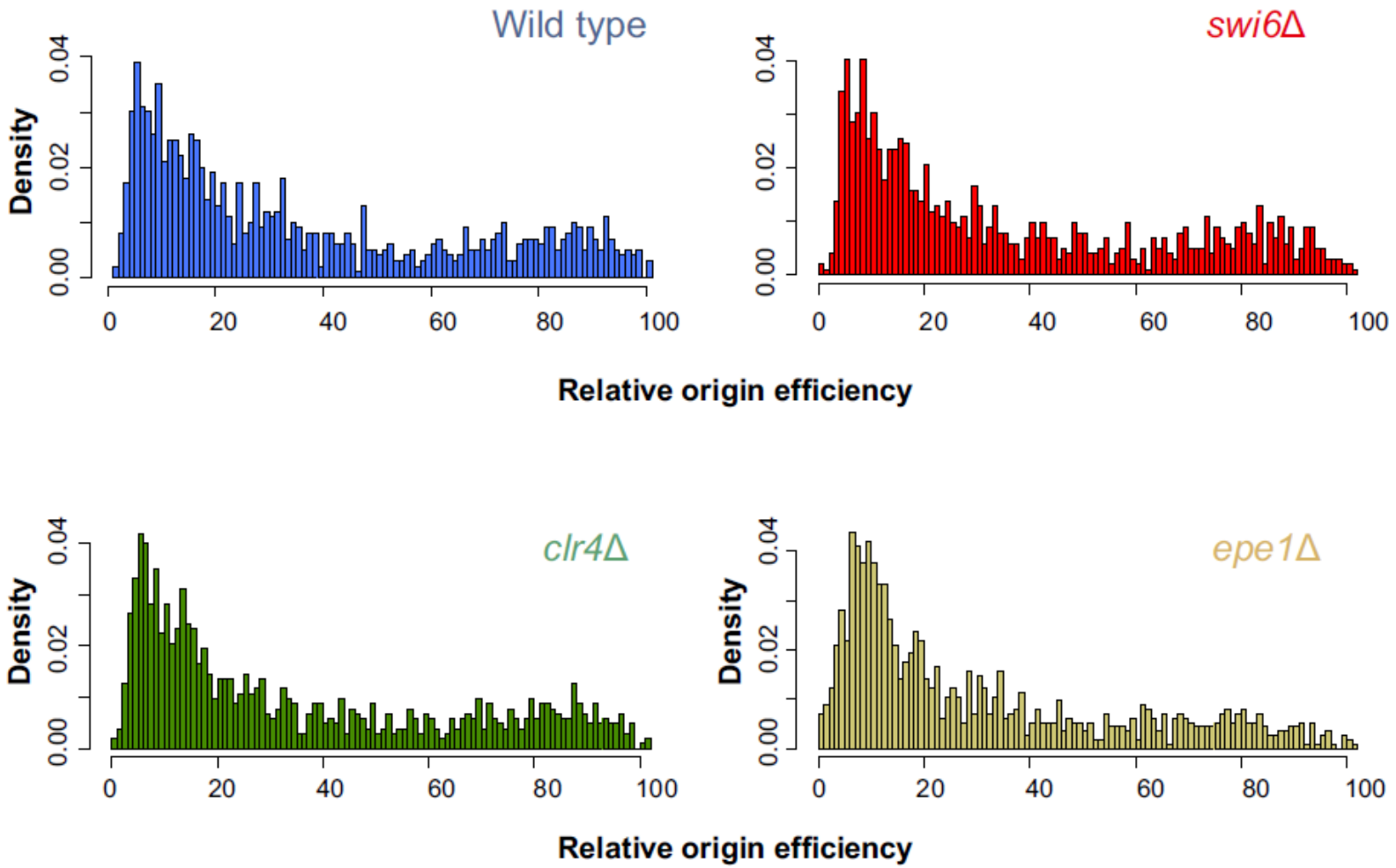
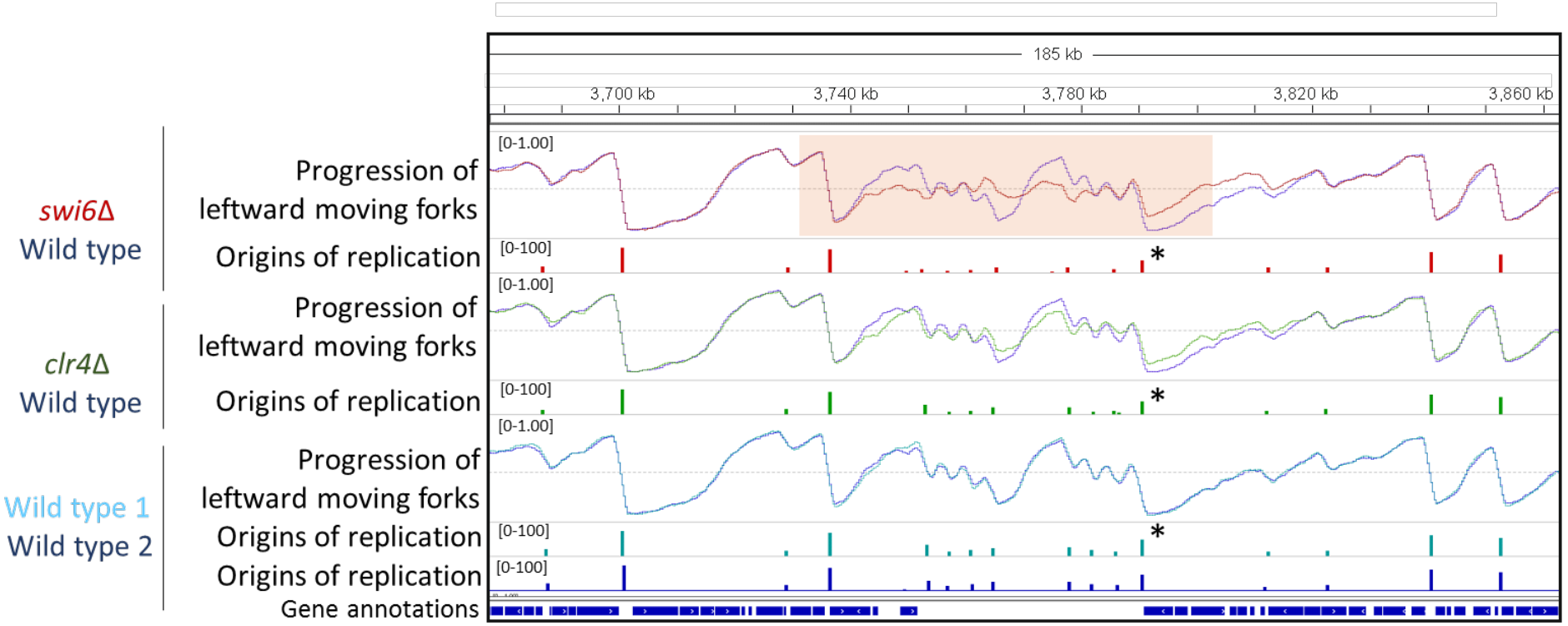

Wild type
swi6Δ
clr4Δ
epe1Δ
Density
0.00
0.02
0.04
0
20
40
60
80
100
Relative origin efficiency

Fig. 4.1 – **Density distribution of origin firing efficiencies in wild type, *swi6*Δ, *clr4*Δ and *epe1*Δ backgrounds**

Origins of replication were mapped for each strains using Pu-Seq, applying a threshold of 0.3, i.e., positive peaks of the differentials of polymerase usage whose heights were above the 30th percentile were mapped as origins. The efficiency of firing was normalised to the value of the 99th percentile in each data set.

A) Replication fork progression



B) Bias towards polymerase δ usage

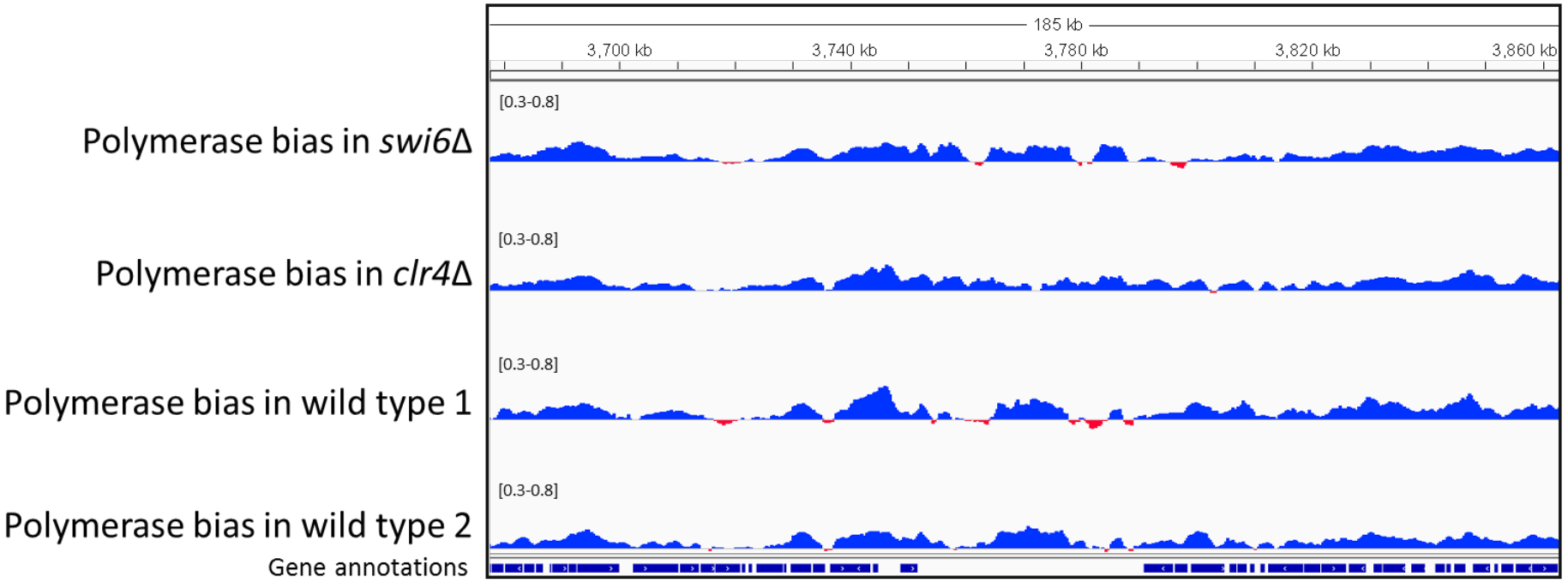


Fig. 4.2 – **Landscape of origin usage, the progression of leftward moving forks and bias of polymerase usage around Centromere 1 in *swi6*Δ, *clr4*Δ and wild type backgrounds**

A) The origins of replication and the progression of leftward moving forks.

Origins of replication were mapped for each strains using Pu-Seq, applying a threshold of 0.3, i.e., positive peaks of the differentials of polymerase usage whose heights were above the 30th percentile were mapped as origins. The efficiency of firing was normalised to the value of the 99th percentile in each data set.

The progression of leftward moving forks (i.e., the average of the usage of polymerase δ on the forward strand and polymerase ε on the reverse strand) was calculated for each background (two independent biological repeats shown in dark and light blue). Deviations from wild type fork progression in the *swi6*Δ are marked in orange.

B) Bias of polymerase usage. The bias of polymerase δ usage on both strands (contribution of polymerase δ at each site / 2) is shown in the range 0.3 - 0.8. Values > 0.5 and < 0.5 indicate a bias towards polymerase δ usage (shown in blue) and or polymerase ε usage (shown in red), respectively. Data from two independent wild type biological repeats are also shown.

4.2.1.2 Replication dynamics *epe1*Δ

The progression of leftward moving forks at Centromere 1 was affected in *epe1*Δ (Fig. 4.3). Similarly to *clr4*Δ and *swi6*Δ, fewer leftward moving forks originating from within the centromere replicated the left-hand portion of the centromere in *epe1*Δ (marked in orange on Fig. 4.3). A similar effect, however, was not seen on the right-hand portion of the centromere. It is possible that the perturbation was the result of an overall less smooth fork progression (seen on the regions flanking the centromere on Fig. 4.3).

In *epe1*Δ cells, the number of origins mapped was outside the wild type range (1146 origins of replication). It is possible that the increase in heterochromatic marks across the genome (Zofall, Yamanaka et al. 2012) acted as a direct or indirect form of replication stress, resulting in the firing of dormant origins. The distribution of origin firing efficiencies remained bimodal (Fig. 4.1), suggesting that the newly firing dormant origins were relatively inefficient.

Polymerase bias around the Centromere 1 in *epe1*Δ did not result in a noticeably different bias towards the usage of polymerase δ than any of the other backgrounds shown in Fig. 4.2 B. The bias, therefore, was too variable to draw meaningful conclusions about the source.

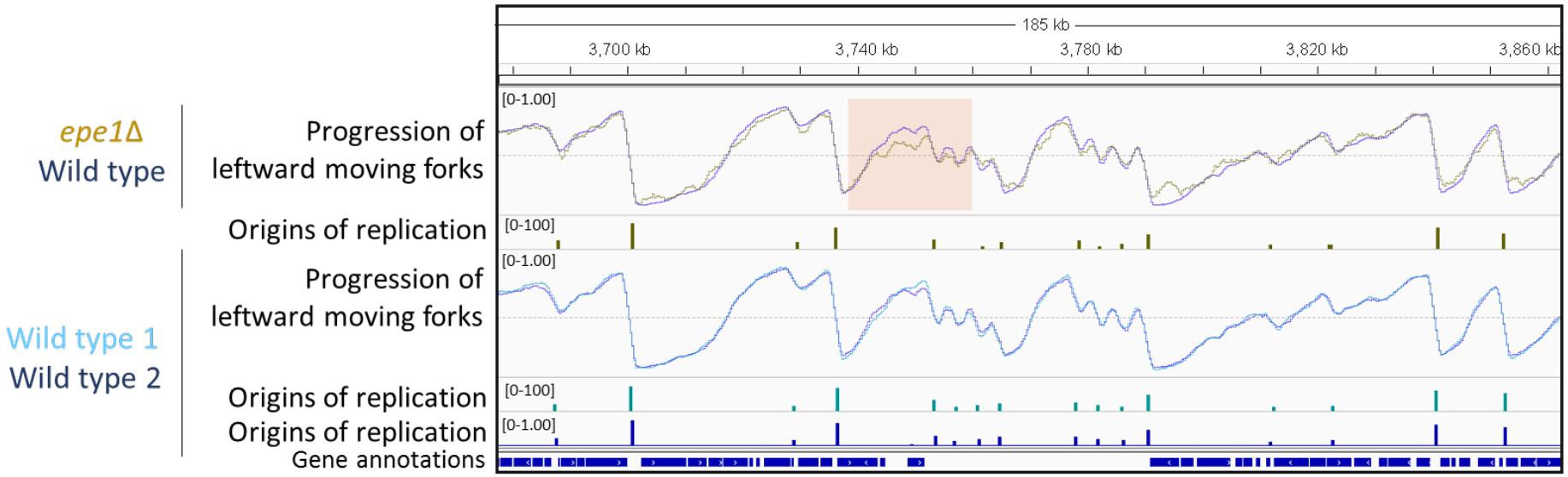


Fig. 4.3 – **Landscape of origin usage and the progression of leftward moving forks around Centromere 1 in *epe1*Δ and wild type backgrounds**

Origins of replication were mapped for each strains using Pu-Seq, applying a threshold of 0.3, i.e., positive peaks of the differentials of polymerase usage whose heights were above the 30th percentile were mapped as origins. The efficiency of firing was normalised to the value of the 99th percentile in each data set.

The progression of leftward moving forks (i.e., the average of the usage of polymerase δ on the forward strand and polymerase ε on the reverse strand) was calculated for each background (two independent biological repeats shown in dark and light blue). Deviations from wild type fork progression in *epe1*Δ are marked in orange and discussed in the text.

4.2.2 Changes to replication dynamics in facultative heterochromatin

Next, we compared the effects that abolishing heterochromatin genome wide (*swi6*Δ and *clr4*Δ) and disrupting it locally (*rif1*Δ and *taz1*Δ) had on replication dynamics around facultative heterochromatin islands.

Given that the 21 facultative heterochromatin islands are not all replicated in the same direction we did not use changes in fork progression as a readout of replication dynamics, so as not to confound the data with directionality. Also, to ensure that all changes in origin usage (even those below the standard 30th percentile threshold) were being taken into account, we did not analyse the normalised origin firing efficiencies. Instead, the differentials of polymerase usage were used directly as a measure of origin activity.

Origin activity around Taz1 dependent and independent facultative heterochromatin islands remained unchanged in *swi6*Δ and *clr4*Δ (Fig. 4.4). A small increase in origin activity was noted in *epe1*Δ. This, however, was likely due to the greater overall number of origins firing in *epe1*Δ and not an effect of the islands themselves.

A substantial increase in origin activity around the 6 Taz1 dependent heterochromatin islands was observed in *rif1*Δ and *taz1*Δ, with the effect being greater in *rif1*Δ than *taz1*Δ (Fig. 4.5 A). The number of origins mapped in *rif1*Δ and *taz1*Δ was within the wild type range (1108 and 1064 origins of replication for *taz1*Δ and *rif1*Δ, respectively - average from 2 independent biological repeats) and there were no noticeable changes in the distribution of origin firing efficiencies (Fig. 4.6). This suggests that the effect was more likely due to an increase in the activity of the origins around the Taz1 heterochromatin islands, rather than the density of the origins in the region. The wild type origin activity around the facultative heterochromatin islands in *swi6*Δ and *clr4*Δ (Fig. 4.4) suggests that the increased origin activity in *rif1*Δ and *taz1*Δ at the Taz1 dependent islands (Fig. 4.5 A) was heterochromatin independent.

No change in origin activity was seen around the 15 Taz1 independent heterochromatin islands (Fig. 4.5 B) or subtelomeres (Fig. 4.7) in *rif1*Δ or *taz1*Δ.

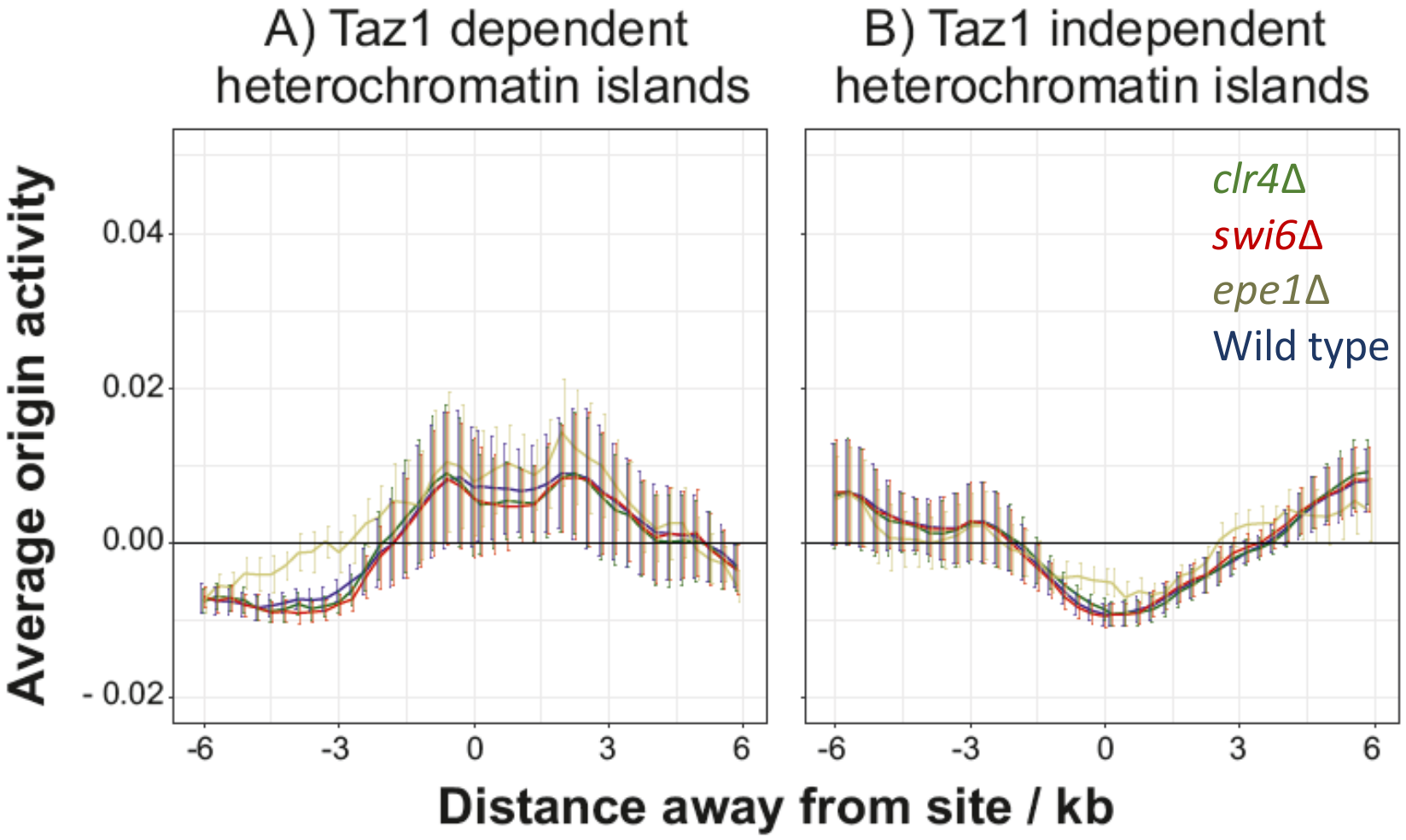

A) Taz1 dependent heterochromatin islands
B) Taz1 independent heterochromatin islands
clr4Δ
swi6Δ
epe1Δ
Wild type
Average origin activity
0.04
0.02
0.00
- 0.02
-6
-3
0
3
6
Distance away from site / kb

Fig. 4.4 – **The average origin activity around Taz1 dependent and independent facultative heterochromatin islands in *clr4*Δ, *swi6*Δ, *epe1*Δ and wild type backgrounds**

The mean origin activity (i.e., the average of the differentials of the polymerase δ usage on the reverse strand and polymerase ε forward strand) in each 300 bp bin is shown for 6 kb up and downstream of the midpoint of the heterochromatin island. The error bars represent the standard error of the mean in each bin. The origin activity shown here were calculated from two independent biological repeats for the wild type data only.

A) The mean activity in each bin averaged across the 6 Taz1 dependent heterochromatin islands.

B) The mean activity in each bin averaged across the 15 Taz1 independent heterochromatin islands.

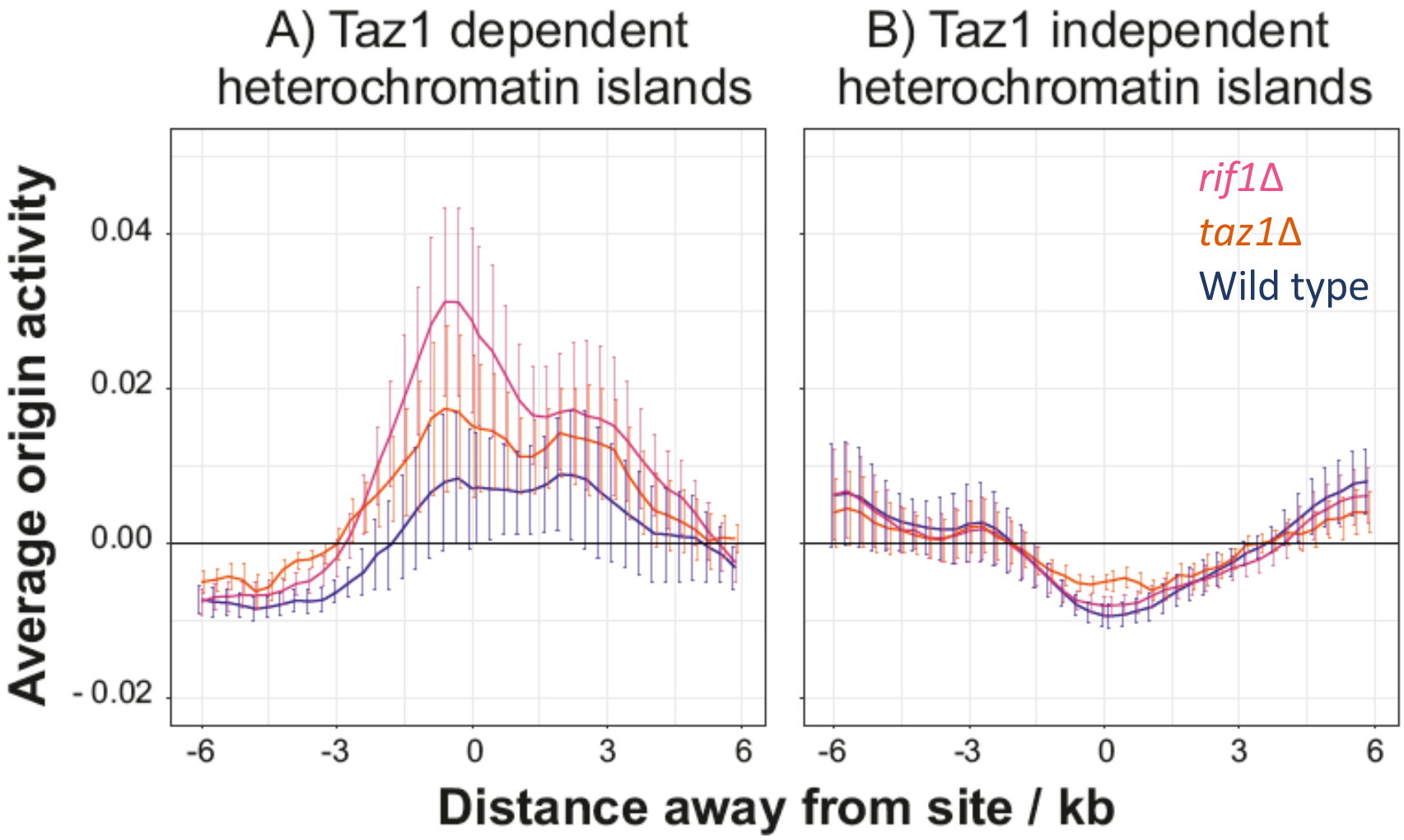

A) Taz1 dependent heterochromatin islands
B) Taz1 independent heterochromatin islands
rif1Δ
taz1Δ
Wild type
Average origin activity
0.04
0.02
0.00
- 0.02
-6
-3
0
3
6
Distance away from site / kb

Fig. 4.5 – **The average origin activity around Taz1 dependent and independent facultative heterochromatin islands in *rif1*Δ, *taz1*Δ and wild type backgrounds**

The mean origin activity (i.e., the average of the differentials of the polymerase δ usage on the reverse strand and polymerase ε forward strand) in each 300 bp bin is shown for 6 kb up and downstream of the midpoint of the heterochromatin island. The error bars represent the standard error of the mean in each bin. The origin activity shown here were calculated from two independent biological repeats.

A) The mean activity in each bin averaged across the 6 Taz1 dependent heterochromatin islands.

B) The mean activity in each bin averaged across the 15 Taz1 independent heterochromatin islands.

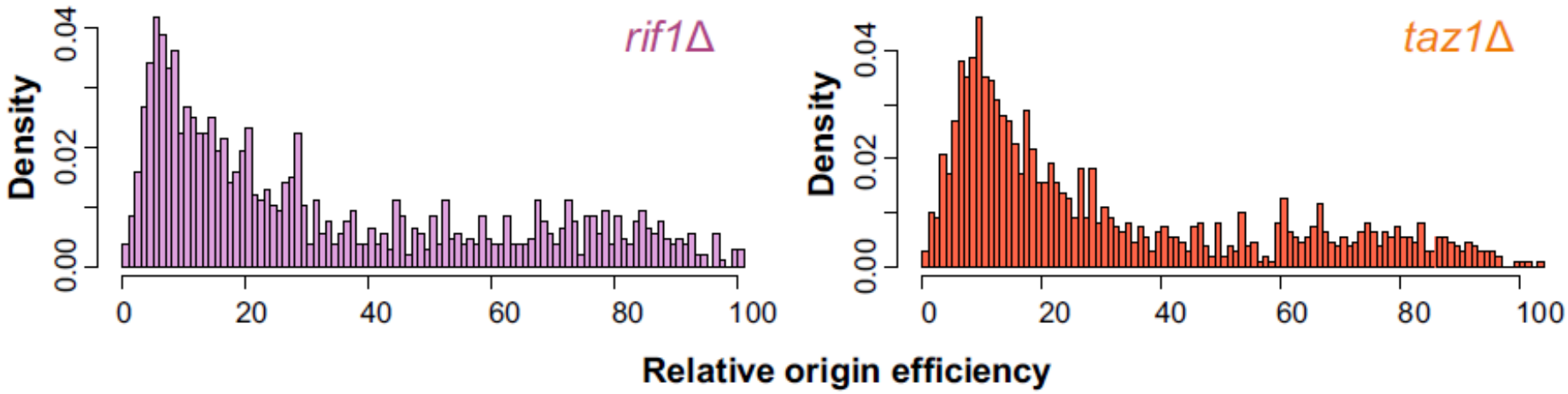


Fig. 4.6 – **Density distribution of origin firing efficiencies in *rif1*Δ and *taz1*Δ backgrounds**

Origins of replication were mapped for each strains using Pu-Seq, applying a threshold of 0.3, i.e., positive peaks of the differentials of polymerase usage whose heights were above the 30th percentile were mapped as origins. The efficiency of firing was normalised to the value of the 99th percentile in each data set.

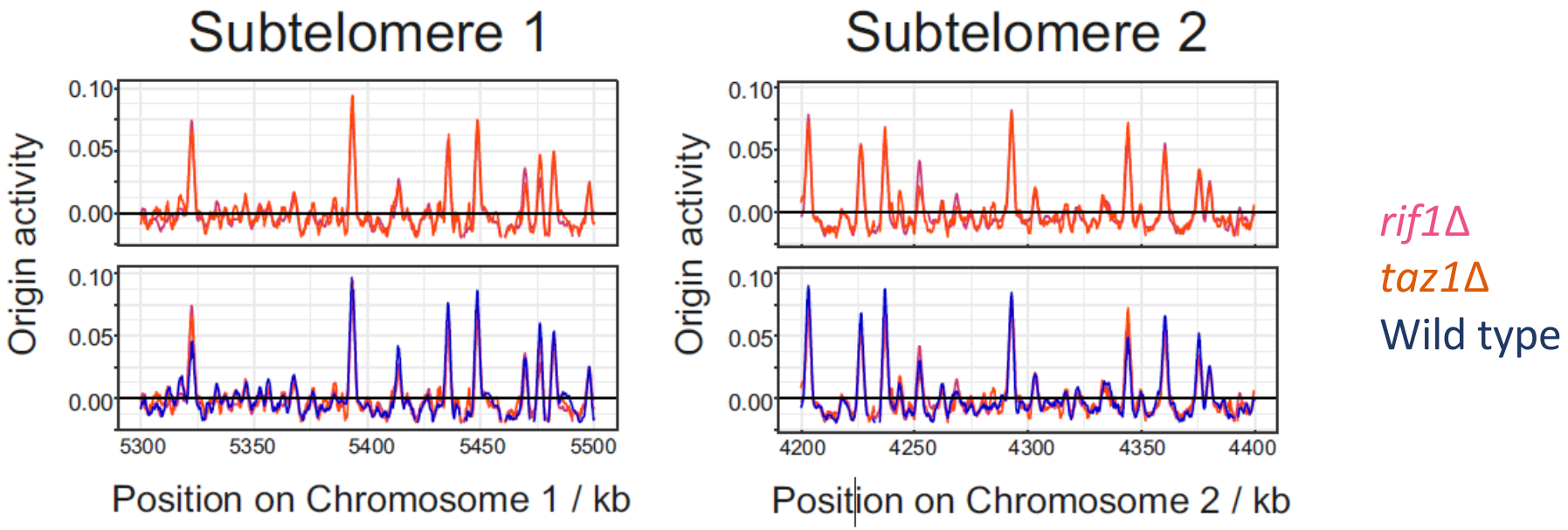


Fig. 4.7 – **The origin activity around subtelomeres in *rif1*Δ, *taz1*Δ and wild type backgrounds**

The mean origin activity (i.e., the average of the differentials of the polymerase δ usage on the reverse strand and polymerase ε forward strand) in each 300 bp bin is shown for a 200 kb regions of the right subtelomeres on Chromosomes 1 and 2.

4.2.3 Effects of *taz1*Δ and *rif1*Δ on replication dynamics

The increased origin activity, shown in Fig. 4.5 A, is in line with previous reports on Rif1 and Taz1 activity. Both Rif1 and Taz1 have been shown to affect origin firing in *S. pombe* (Hayano, Kanoh et al. 2012, Tazumi, Fukuura et al. 2012). It is also known that the two proteins interact and Taz1 recruits Rif1 to telomeres (Kanoh and Ishikawa 2001). It is, therefore, possible that the inhibition of origin activity was caused by Rif1 or Taz1 alone, with the other acting as a recruiter to the binding site.

To determine whether the increase in origin activity described in 4.2.2 was entirely due to the loss of Rif1 and/or Taz1 binding, we analysed the changes in origin activity and local RT around other reported Rif1 binding sites (Rif1 BSs). No sites other than subtelomeres and the Taz1 dependent heterochromatin islands have been reported as binding sites for Taz1.

4.2.3.1 Effect of *taz1*Δ and *rif1*Δ on origin firing

A ChIP-Seq analysis of Rif1-6×His-10×Flag in G_1, in an *nda3-KM311* background reported 90 Rif1 BSs, from which a consensus sequence for a Rif1 binding motif was extracted (Kanoh, Matsumoto et al. 2015). 35 out of the 90 reported sites contained two or more repeats of this motif and these 35 sites were described as "strong" Rif1 BSs. Rif1 binding to these sites was seen in *taz1*Δ, suggesting that, unlike at Taz1 dependent facultative heterochromatin islands, Rif1 binding to strong Rif1 BSs was not Taz1 dependent. Interestingly, Taz1 dependent islands were not part of the Rif1 BSs reported by (Kanoh, Matsumoto et al. 2015), even in the presence of Taz1.

Given the ability of strong Rif1 BSs to adopt G-quadruplex structures *in vitro* and the preferential binding of G-quadruplex DNA by purified Rif1 protein, it was hypothesised that in *S. pombe* Rif1 binds G-quadruplexes *in vivo* (Kanoh, Matsumoto et al. 2015). We, therefore, looked at the effect of *rif1*Δ and *taz1*Δ at strong Rif1 BSs and, independently mapped, G-quadruplexes (Sabouri, Capra et al. 2014).

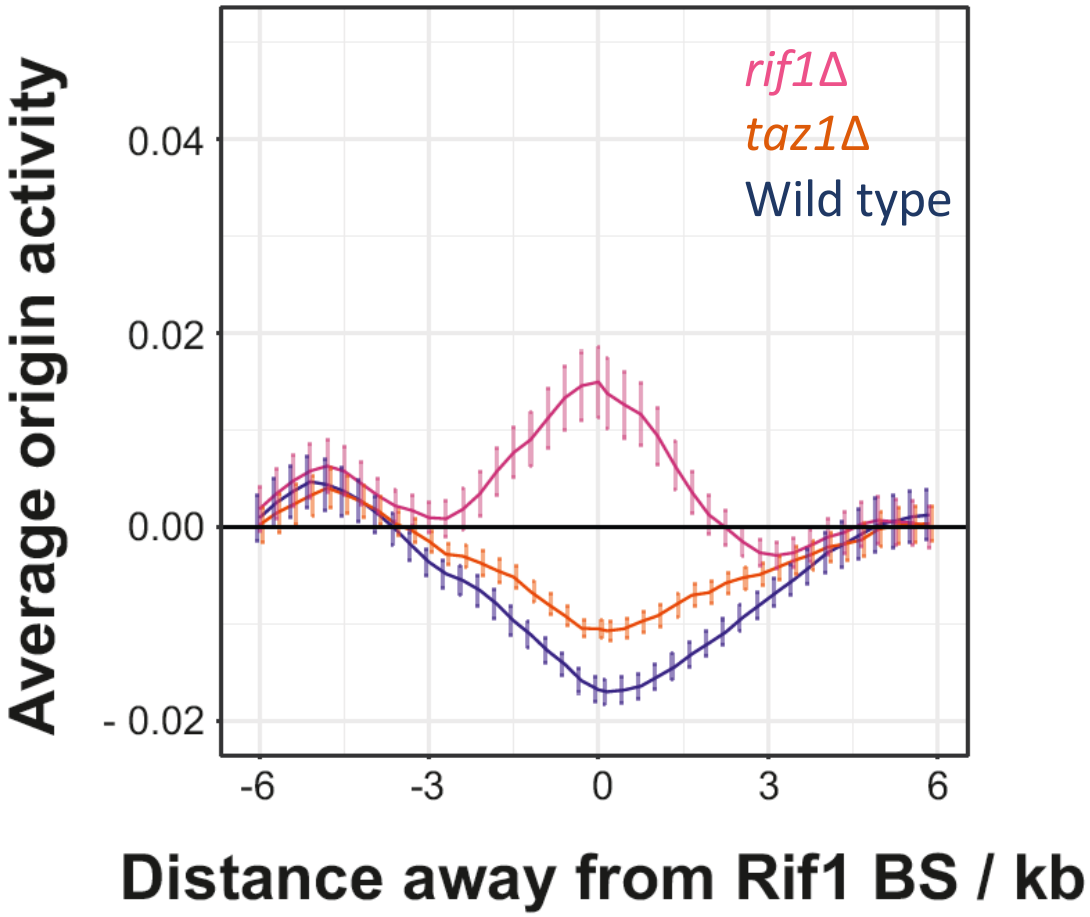
In wild type *S. pombe*, strong Rif1 BSs appear to be regions where few origins fire and replication forks merge, i.e., they are potential termination sites (evidenced by the negative value of the differential of polymerase usage) - Fig. 4.8 A. In *rif1*Δ, we

observed an increase in origin firing up to 3 kb either side of the BSs, suggesting that Rif1 binding locally inhibits origin firing. We did not see evidence of long-range effects (>50 kb) on the origin activity (previously reported in (Kanoh, Matsumoto et al. 2015)). We also saw a marginal increase in origin activity in *taz1*Δ - Fig. 4.8 A.

Around G-quadruplex sites, we saw a very small, but reproducible, increase in origin activity in *rif1*Δ (Fig. 4.8 Bi). To determine whether the increased origin activity at a small number of strong Rif1 BSs caused this effect, we analysed the G-quadruplex data where the positions that overlapped strong Rif1 BSs were removed. Only 6 (out of 446) of the mapped G-quadruplex positions overlapped with the 35 strong Rif1 BSs. The very small increase in origin activity was maintained in *rif1*Δ around these 440 G-quadruplex positions (Fig. 4.8 Bii). This suggests that the effect shown in Fig. 4.8 Bi was likely to have been caused by an interaction of Rif1 with G-quadruplex DNA. The fact, however, that this effect is so small and that only 6 out of the 446 G-quadruplex sites overlap with strong Rif1 BSs suggest that despite the ability of purified Rif1 to bind G-quadruplex DNA *in vitro*, it may not have a functional role *in vivo*. No change in origin activity was observed in *taz1*Δ around the G-quadruplexes (Fig. 4.8 B).

These data also show that the action of Rif1, not Taz1, inhibits origin firing locally. This implies that the increased origin activity shown in Fig. 4.5 A was due to the loss of Rif1 at these sites- either directly in *rif1*Δ or due to an abrogated binding of Rif1 to these sites in *taz1*Δ. The increase in origin firing was not as high in *taz1*Δ compared to *rif1*Δ, suggesting that Rif1 may be able to bind inefficiently to Taz1 dependent heterochromatin islands in a Taz1 independent manner. The data also suggest that Taz1 independent heterochromatin islands may be a subset of Rif1 BSs, despite not being identified as Rif1 BSs by (Kanoh, Matsumoto et al. 2015).

A) Origin activity around Rif1 binding sites



B) Origin activity around G-quadruplexes

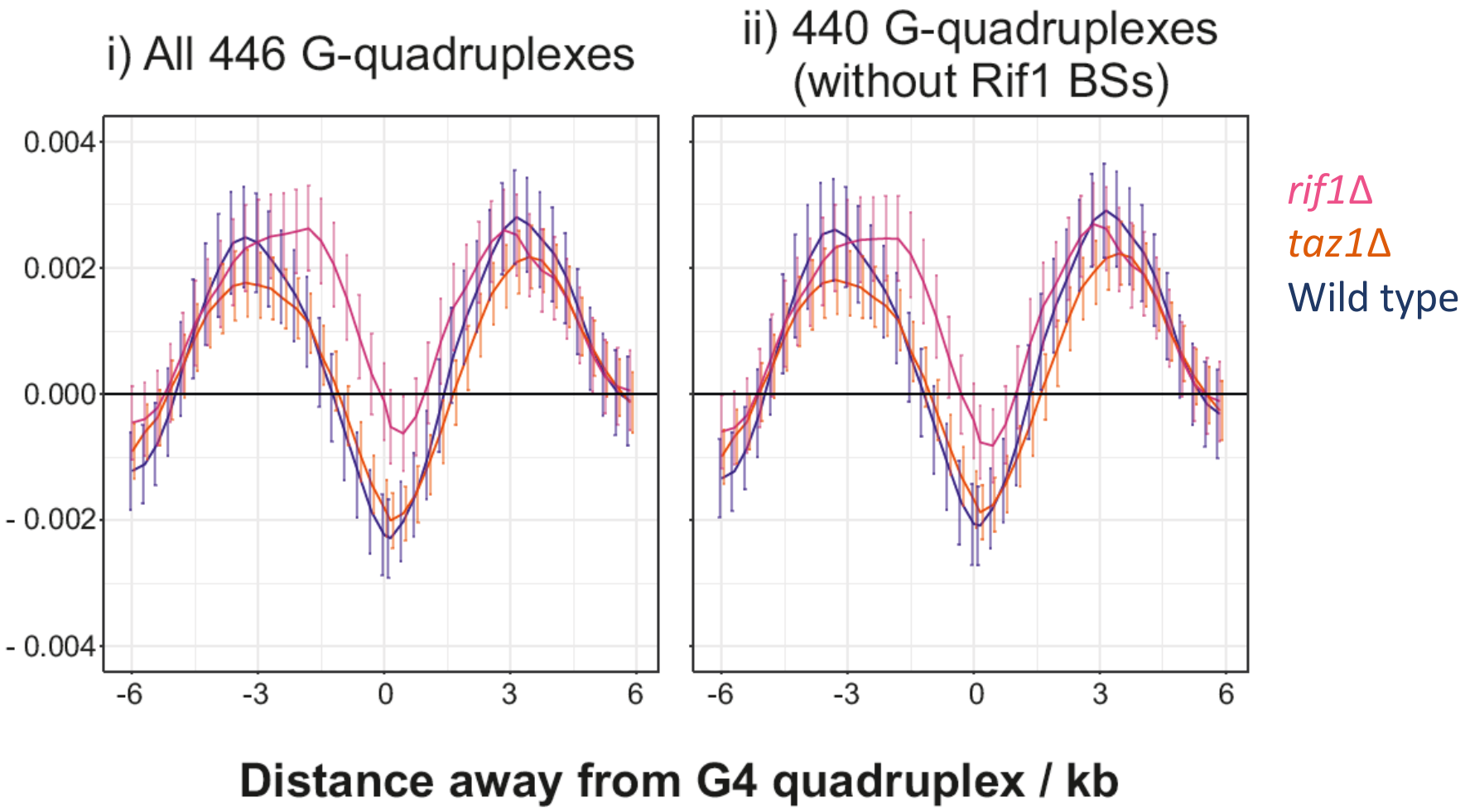


Fig. 4.8 – **The average origin activity around strong Rif1 binding sites G quadruplexes in *rif1*Δ, *taz1*Δ and wild type backgrounds**

The mean origin activity (i.e., the average of the differentials of the polymerase δ usage on the reverse strand and polymerase ε forward strand) in each 300 bp bin is shown for 6 kb up and downstream of the midpoint of the heterochromatin island. The error bars represent the standard error of the mean in each bin. The origin activity shown here were calculated from two independent biological repeats for the wild type data only.

A) The mean activity in each bin averaged across the 35 strong Rif1 binding sites.

B) The mean activity in each bin averaged across G-quadruplexes. The mean activity was counted across all of the 446 G-quadruplexes (i) and the 440 G-quadruplexes that did not overlap with any strong Rif1 binding sites (ii).

4.2.3.2 Effect of *taz1*Δ and *rif1*Δ on replication timing

Historically, data describing origin firing and replication timing was derived from BrdU incorporation assays. Briefly, *S. pombe* cultures were synchronised (often using cell cycle mutants) and released into BrdU in the presence of HU, which blocks cells in early S-phase, ensuring the firing of only early origins. BrdU, an analogue of thymidine, is incorporated into the DNA, acting as a marker of nascent DNA. Sonicated DNA fragments containing BrdU were be pulled down using an anti-BrdU antibody. The nascent DNA was then either hybridised to a tiling array or sequenced, giving an estimate of the copy number of each locus (Ryba, Battaglia et al. 2011).

BrdU incorporation techniques reliably identify efficiently firing early origins. They are, however, unable to differentiate between origins which fire early and those which fire efficiently, if those are fired relatively late in S-phase. Additionally, the commonly used mutations for synchronising *S. pombe* have been reported to have affect origin usage (Xu, Yanagisawa et al. 2012). The application of BrdU incorporation assays and cell cycle mutants to study origin firing patterns can, therefore, lead to false conclusions.

As described in 2.2.4.3.2, Pu-Seq data can be analysed to generate an RT profile (Pu-Seq Trep) of each chromosome. These patterns are not relative, i.e., the data cannot be used to compare changes between chromosomes. To determine how the changes in origin firing efficiencies around Taz1 dependent heterochromatin islands and the Rif1 BSs affected the local replication timing around them, we analysed the Pu-Seq Trep around those regions in *taz1*Δ and *rif1*Δ.

4.2.3.2.1 Changes seen in *taz1*Δ

The global RT program in *S. pombe* is biphasic (Daigaku, Keszthelyi et al. 2015). Chromosome 3, the left arm of Chromosome 2 and small parts on the right arm of Chromosome 1 replicate early in S-phase (marked in green on Fig. 4.9), with the rest of the genome replicating later. This pattern is most visible on Chromosome 2 – replication timing transitions between early and late midway through the chromosome (to the right on Centromere 2).

The *taz1*Δ global RT profile was comparable to that of wild type - Fig. 4.10 A. All changes in replication timing were limited to Taz1 dependent heterochromatin islands and subtelomeres.

Replication timing changes around the Taz1 dependent heterochromatin islands were very pronounced in *taz1*Δ (Fig. 4.10 B, C and D). On average, across all 6 Taz1 dependent heterochromatin islands, the relative replication timing spiked around the site in *taz1*Δ and levelled off to wild type levels between 40-50 kb away from the site (Fig. 4.11 A). This suggests that local changes to origin firing (Fig. 4.5 A) affect the regional replication timing.

Additionally, changes in Pu-Seq Trep around the island on the right arm of Chromosome 1 resulted in a relative replication timing change of an adjacent replication timing domain. In *taz1*Δ, the earlier replication of the region adjacent to the island (marked in green on Fig. 4.10 B ii), had a direct effect on decreasing the regional replication timing of the ~200 kb domain next to it (marked in red on Fig. 4.10 B ii). It is unclear why this effect of compensating replication timing was not seen at the other Taz1 dependent heterochromatin islands.

A modest increase in local replication timing around the Rif1 BSs was noted in *taz1*Δ (Fig. 4.11 B), correlating well with the small increase in the origin activity (Fig. 4.8 A).

Subtelomeres in *taz1*Δ were replicated earlier than wild type, with the exception of the right subtelomere of Chromosome 3, which replicated later (Fig. 4.10 A). The late replication timing phenotype on Chromosome 3 span approximately 0.5 Mb proximal to the right telomere. Given that both ends of Chromosome 3 are "capped" by repetitive rDNA units (Uzawa and Yanagida 1992), it is unclear why they would have different phenotypes in *taz1*Δ. Given the long range effect, however, it is unlikely that this change is caused by anomalous read alignment in the subtelomeric regions. Surprisingly, however, the origin activity around right subtelomeric regions of Chromosomes 1 and 2 was almost identical to that of wild type (Fig. 4.7), suggesting that the late replicating phenotype was not due to the inhibition of origin firing activity in the region.

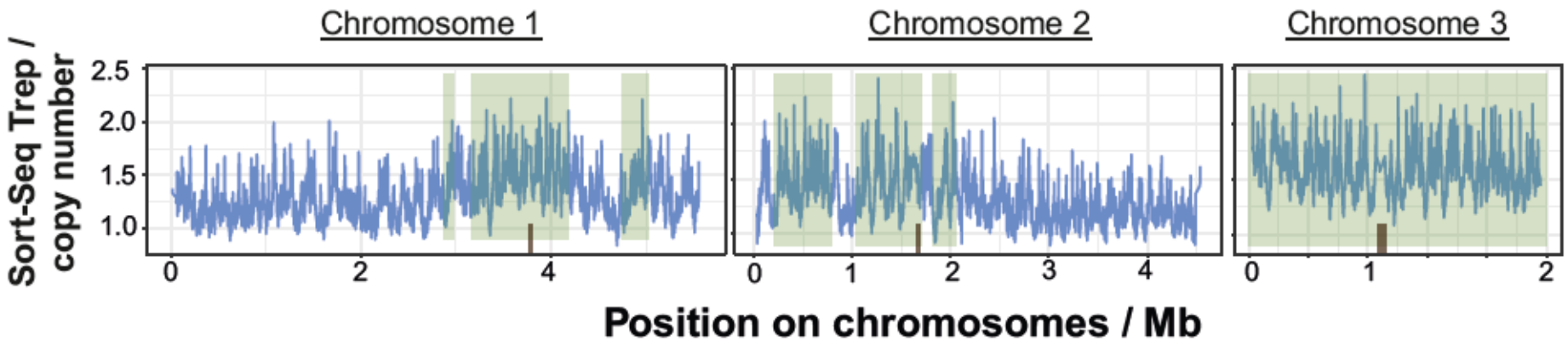
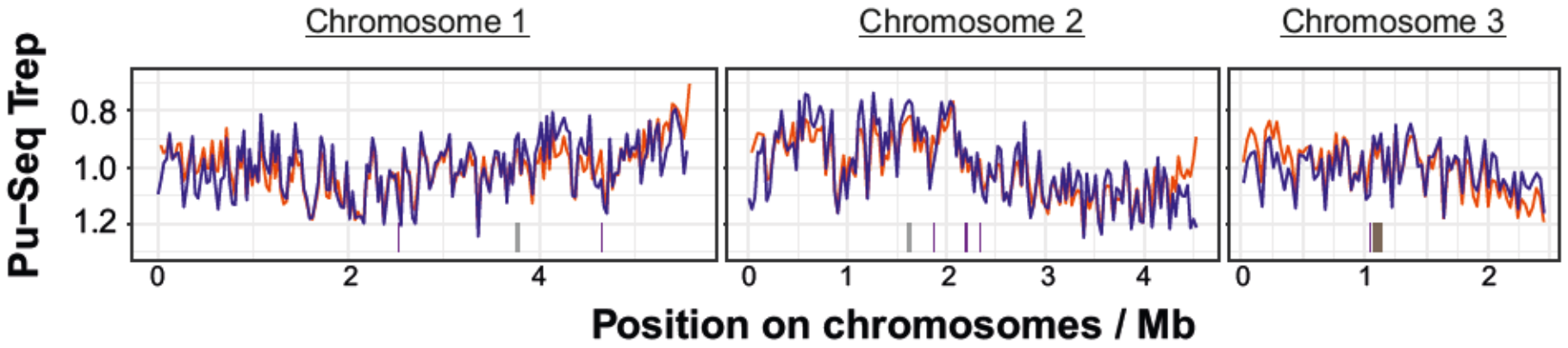


Fig. 4.9 – **Sort-Seq Trep across all three chromosomes in wild type *S. pombe***

Cells were sorted by FACS from an asynchronous culture and sequencing data from synchronous S-phase and G_2 populations were compared to produce a Sort-Seq copy number profile. The data were collected, analysed published in (Daigaku, Keszthelyi et al. 2015).

A) Pu-Seq Trep genome wide

taz1Δ
Wild type



B-D) Pu-Seq Trep around Taz1 islands

taz1Δ
Wild type

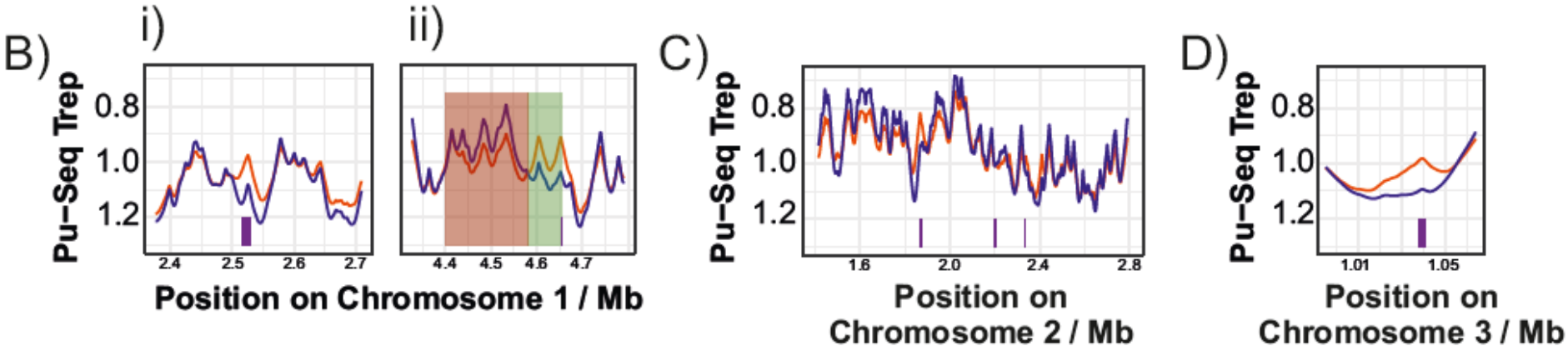


Fig. 4.10 – **Pu-Seq Trep across all three chromosomes and around Taz1 dependent facultative heterochromatin islands in *taz1*Δ and wild type backgrounds**

A) Pu-Seq Trep across all three chromosomes in *taz1*Δ and wild type. Pu-Seq Trep was calculated using the progression of leftward moving forks and assuming a constant fork velocity of 1.5 kb/min. It represents when in S-phase each locus is replicated. The y-axis was, therefore, inverted to maintain the convention of early and late replicating regions being shown on the top and bottom of the graph, respectively. Centromeres and Taz1 dependent facultative heterochromatin islands are marked in grey and purple, respectively. The wild type Pu-Seq Trep profile is derived from an average of data generated by 5 wild type Pu-Seq experiments.

B-D) Pu-Seq Trep across each of the 6 Taz1 dependent facultative heterochromatin islands. Pu-Seq Trep was calculated as explained in A. Late to early and early to late changes between wild type and *taz1*Δ are shown in red and green, respectively. Centromeres and Taz1 dependent facultative heterochromatin islands are marked in grey and purple, respectively.

Figures B i and Bii show the relative Pu-Seq Trep around two different Taz1 dependent heterochromatin islands on the left and right arms of chromosome 1, respectively. Both heterochromatin islands are shown on chromosome 1 on Fig. 4.10 A.

The wild type Pu-Seq Trep profile is derived from an average of data generated by 5 wild type Pu-Seq experiments.

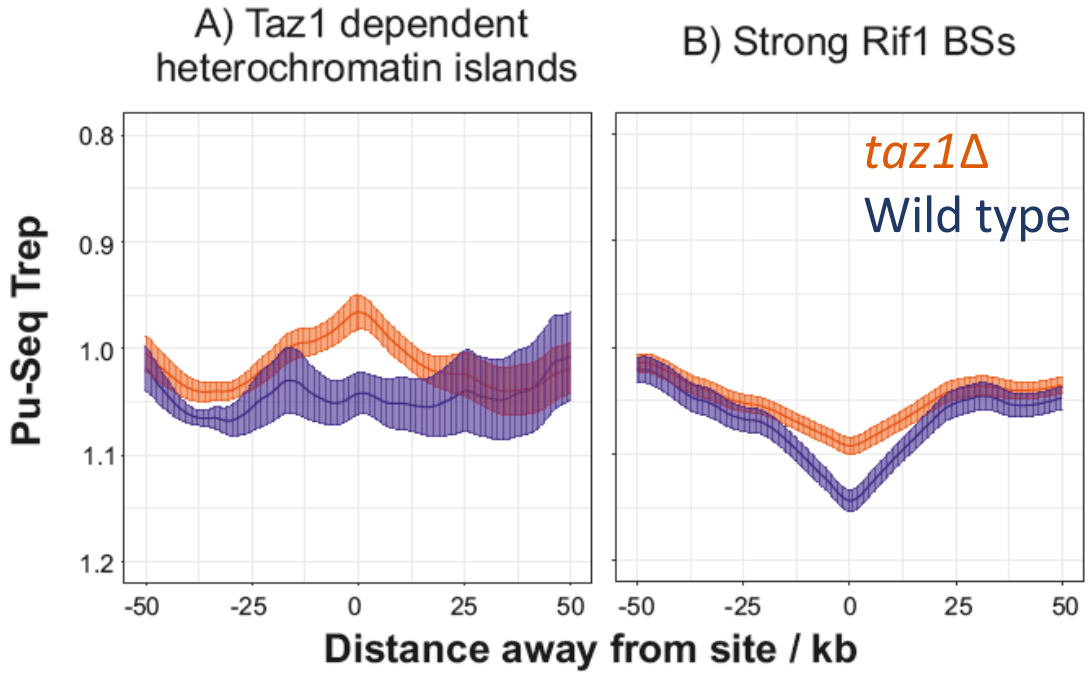


Fig. 4.11 – **Pu-Seq Trep around Taz1 dependent facultative heterochromatin islands and strong Rif1 binding sites in *taz1Δ* and wild type background**

Pu-Seq Trep was calculated using the progression of leftward moving forks and assuming a constant fork velocity of 1.5 kb/min. It represents when in S-phase each locus is replicated. The y-axis was, therefore, inverted to maintain the convention of early and late replicating regions being shown on the top and bottom of the graph, respectively. Pu-Seq Trep 6 kb up and downstream of the midpoint of each site was recorded. The error bars represent the standard error of the mean in each bin. The Pu-Seq Trep shown here were calculated from two independent biological repeats for both backgrounds.

A) The Pu-Seq Trep in each bin averaged across the 6 Taz1 dependent heterochromatin islands.

B) The Pu-Seq Trep in each bin averaged across the 35 strong Rif1 binding sites.

4.2.3.2.2 Changes seen in *rif1*Δ

Unlike in *taz1*Δ, replication timing changes in *rif1*Δ were not limited to subtelomeres and Taz1 dependent heterochromatin islands. The global program of replication timing was affected in *rif1*Δ - Fig. 4.12.

Similarly to *taz1*Δ, an increase in the relative RT was noted around Taz1 dependent heterochromatin islands and Rif1 BSs in *rif1*Δ (Figs. 4.13 A and B). This increase in Pu-Seq Trep was greater than that seen in *taz1*Δ but did not span regions as large - the increase in relative replication timing levelled off after ~10 kb. Interestingly, in *rif1*Δ, subtelomeres also replicated earlier than in wild type (Fig. 4.12), without a concomitant increase in the origin activity in the region (Fig. 4.7).

While local/ regional changes in replication timing around Rif1 BSs and Taz1 dependent heterochromatin islands in *taz1*Δ and *rif1*Δ can be explained by the inhibition of origin firing by Rif1 in those regions (Figs. 4.5 A and 4.8 A), the global replication timing pattern in *rif1*Δ cannot. The global distribution of origin firing efficiencies, and the total number of origins fired, in *rif1*Δ was wild type (Fig. 4.6). Similarly the change in the relative RT of subtelomeres (Figs. 4.10 A and 4.12) does not correlate with changes in origin activity (Fig. 4.7). This suggests that Rif1 may have another role in controlling replication timing in *S. pombe*, which is independent from its origin inhibitory role.

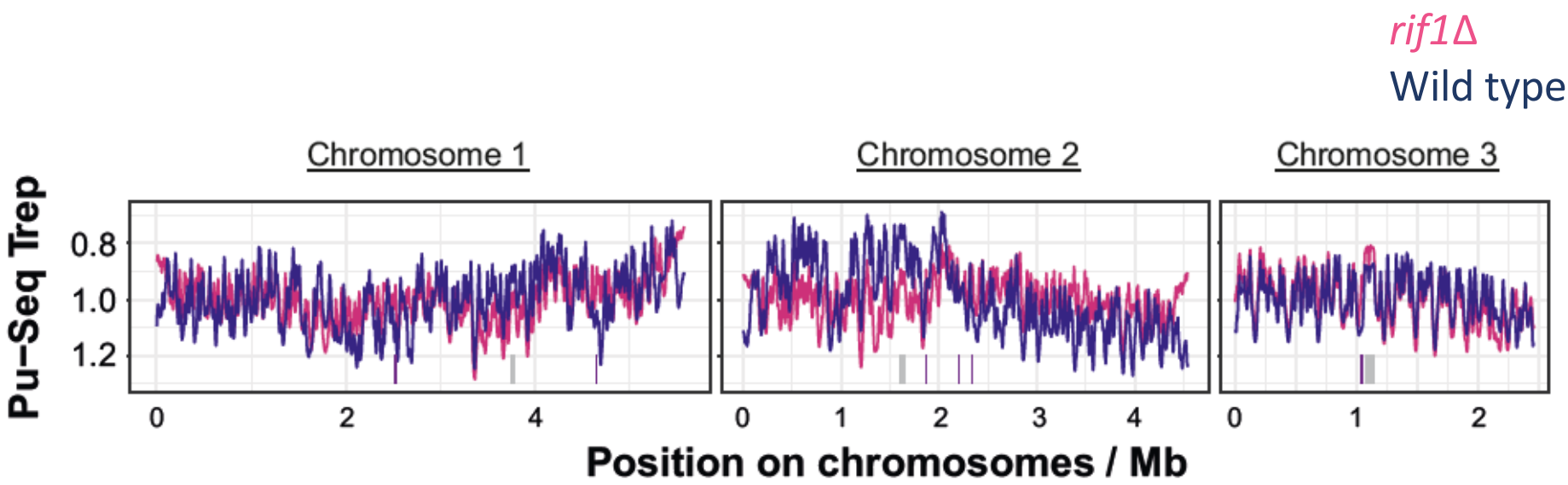


Fig. 4.12 – **Pu-Seq Trep across all three chromosomes in *rif1Δ* and wild type *S. pombe***

Pu-Seq Trep was calculated using the progression of leftward moving forks and assuming a constant fork velocity of 1.5 kb/min. It represents when in S-phase each locus is replicated. The y-axis was, therefore, inverted to maintain the convention of early and late replicating regions being shown on the top and bottom of the graph, respectively. Centromeres and Taz1 dependent facultative heterochromatin islands are marked in grey and purple, respectively. The wild type Pu-Seq Trep profile is derived from an average of data generated by 5 wild type Pu-Seq experiments.

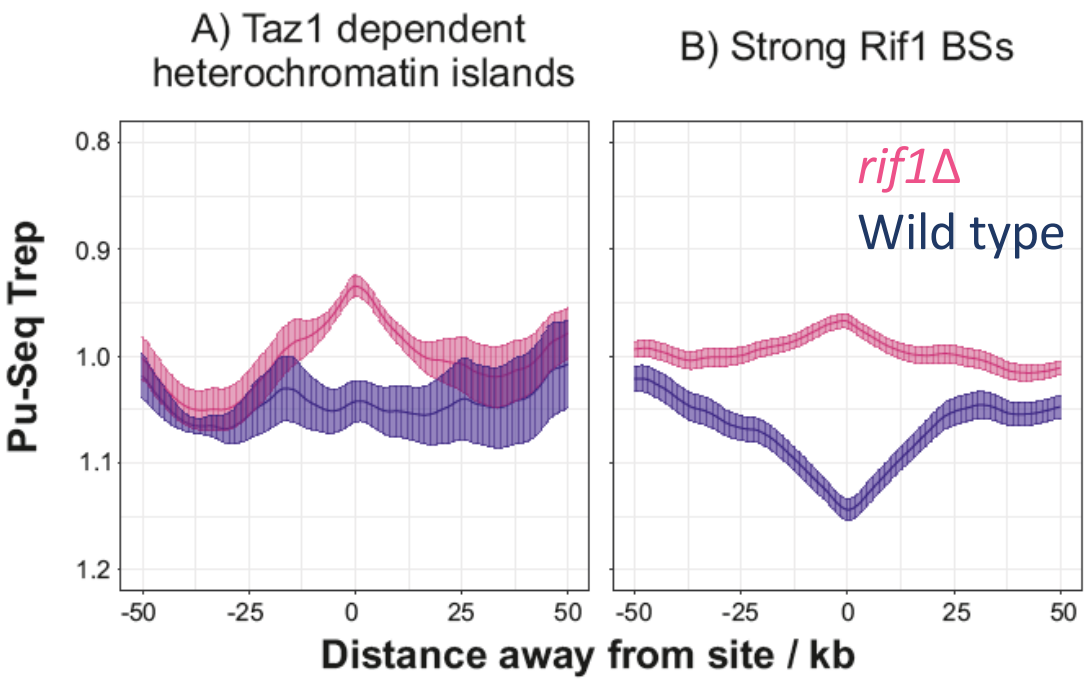


Fig. 4.13 – **Pu-Seq Trep around Taz1 dependent facultative heterochromatin islands and strong Rif1 binding sites in *rif1Δ* and wild type backgrounds**

The Pu-Seq Trep was calculated using the progression of leftward moving forks and assuming a constant fork velocity of 1.5 kb/min. Pu-Seq Trep represents when in S-phase at which each locus is replicated. The y-axis was, therefore inverted to maintain the convention of early and late replicating regions being shown on the top and bottom of the graph, respectively. Pu-Seq Trep 6 kb up and downstream of the midpoint of the site was recorded. The error bars show the standard error of the mean in each bin. The Pu-Seq Trep shown here were calculated from two independent biological repeats for all backgrounds.

A) The Pu-Seq Trep in each bin averaged across the 6 Taz1 dependent heterochromatin islands.

B) The Pu-Seq Trep in each bin averaged across the 35 strong Rif1 binding sites.

4.3 Discussion

The wild type pattern of origin firing around Centromere 1 was affected in *swi6*Δ and *clr4*Δ (Fig. 4.2 A). It is possible that the wild type pattern of origin firing was lost due to a change in the regional replication timing of the centromere. Swi6 interacts with Dfp1, the regulatory subunit of DDK (Bailis, Bernard et al. 2003) and DDK phosphorylation stimulates the loading of Sld3 and results in the early replication of centromeres (Hayashi, Takahashi et al. 2009). In *swi6*Δ, where the early replication phenotype of the centromere is lost (Hayashi, Takahashi et al. 2009), the relative decrease in the efficiency of the origins firing within the centromere could have been caused by an increase in the passive replication of the region. With origins within and around the centromere firing at the same time in S-phase, the probability of the centromere being passively replicated by rightward and leftward moving replication forks originating on the left and right-hand sides of the centromere, respectively, increases. This would have affected the pattern of leftward moving forks as shown in Fig. 4.2 A.

It is also possible, and not mutually exclusive, that the repetitive nature of the centromere in the absence of heterochromatin caused the stalling and collapsing of forks. The perturbed replication shown in Fig. 4.2 A could have been caused, at least in part, by more forks collapsing within the repetitive regions and being rescued by inefficiently firing dormant origins. This interpretation would suggest that on the contrary to acting as an RFB, wild type levels of heterochromatin could be important for the maintenance of replication fidelity in repetitive regions.

The changes in the origin firing and fork progression were less severe in *clr4*Δ, compared to s*wi6*Δ (Fig. 4.2 A). It has been shown, using immune-localisation (Ekwall, Nimmo et al. 1996) and ChIP around Centromere 1 (Partridge, Borgstrom et al. 2000), that Swi6 is recruitment to centromeres is dependent on Clr4. Given that *clr4*Δ replication did not phenocopy that of *swi6*Δ, it is possible that in *clr4*Δ, Swi6 is recruited to centromeres via a redundant and less efficient mechanism and that the low levels of Swi6 are enough to establish wild type heterochromatin and/or early replication timing.

Despite origin activity and fork progression being locally affected at Centromere 1 in the absence of heterochromatin, there was no reproducible bias towards polymerase δ (Fig. 4.2 B). This led us to conclude that, given the current level of noise in the Pu-Seq data, we cannot use polymerase δ bias as an accurate readout of HR restarted at repetitive regions.

An increased load of heterochromatin affected global fork progression and caused an increase in the number of origins fired. The progression of replication forks was not as smooth as that seen in wild type or any other mutant background (Fig. 4.3). It is, therefore, possible that an increase in heterochromatin marks across the genome (as was reported for *epe1*Δ by (Zofall, Yamanaka et al. 2012)) can act as an RFB.

Interestingly, the changes to fork progression in *epe1*Δ around Centromere 1 (Fig. 4.3) were similar to those seen in *clr4*Δ (Fig. 4.2 A). Due to the globally perturbed fork progression in *epe1*Δ, however, it is impossible to say whether this is a centromere specific effect. It could be speculated, however, that although the interaction of Swi6 with Epe1 was not reported to be involved in the ability of Swi6 to recruit Dfp1 the early replication timing around centromeres may be affected to a small extend in *epe1*Δ. It is also possible that the increased heterochromatic marks led to more frequent fork stalling and potentially collapse. The number and distribution of the origins fired around the centromere in *epe1*Δ were comparable to wild type, which does not suggest dormant origin firing. It is, however, possible that the origins were firing below the noise threshold and not being detected. It is important to note that we have not yet carried out an independent biological replicate of the *epe1*Δ Pu-Seq analysis. We cannot, therefore, rule out that the unusual progression of forks in *epe1*Δ was caused by a technical error.

At facultative heterochromatin islands, loss of heterochromatin did not have any effect on origin activity (Fig. 4.4). The deletion of facultative heterochromatin factors, Rif1 and Taz1, however, did have a very drastic effect on origin firing at Taz1 dependent heterochromatin islands. (Fig. 4.5 A). We, therefore, suggest that the origin firing inhibition effect, driven by Rif1 was heterochromatin independent. A similar effect of origin inhibition was seen at previously published Rif1 BSs (Fig.

4.8 A), suggesting that Taz1 dependent heterochromatin islands are indirect binding sites for Rif1.

The changes in origin firing generally correlated with the changes in Pu-Seq Trep. A modest increase in origin activity in *taz1*Δ around Rif1 BSs (Fig 4.8 A) resulted in a small increase in local replication timing around those sites (Fig. 4.10 B). A large increase in origin activity in *rif1*Δ, on the other hand, correlated with a greater increase in local RT around the same sites (Figs. 4.8 A and 4.13 B). A similar effect was seen at Taz1 dependent heterochromatin islands - an increase in origin firing in *taz1*Δ and *rif1*Δ (Fig. 4.5 A) led to a spike in RT that persisted several tens of kb (Figs. 4.11 A and 4.13 A). It is possible that the long range effects of *rif1*Δ reported by (Kanoh, Matsumoto et al. 2015) were long range consequences of delayed timing, not origin inhibition. Interestingly, however, (Kanoh, Matsumoto et al. 2015), analysed changes around Rif1 BSs, not Taz1 islands and the former did not, on average, result in relative RT changes that span >50 kb.

An interesting exception to the straightforward relationship between origin activity and relative RT was the Taz1 dependent island on the right arm of Chromosome 1 (Fig. 4.10 B). The increase in the RT around the Taz1 dependent island was compensated by a change in the regional timing of the adjacent early replicating domain (Fig. 4.10 Bii). We propose that, in wild type cells, the early replication of the larger domain was maintained by the inhibition of origin firing around the Taz1 dependent island. This inhibition allowed the replication forks that originated in the early replicating domain to passively replicate the Taz1 dependent island adjacent to it. In *taz1*Δ, origins fired with equal probabilities in the red and green regions, decreasing the likelihood of the Taz1 dependent island being passively replicated. These changes result in the loss of the two replication timing "domains" across the ~250 kb region in *taz1*Δ.

This effect was not seen at the other Taz1 dependent islands. It is possible, therefore, that this change in timing reflects a genuine biological function that was impeded by *taz1*Δ at that site. As discussed in 4.1.2, the Taz1 dependent facultative heterochromatin islands have been independently discovered in two separate functional contexts (Zofall, Smith et al. 2016, Toteva, Mason et al. 2017). It is possible that they play other roles, which are not yet described.

The change in the relative RT of subtelomeres, on the other hand, in both *rif1*Δ (Fig. 4.12) and *taz1*Δ (Fig. 4.10 A) did not correlate with changes in the origin activity in those regions (Fig. 4.7). Another unexpected result of this analysis was the observation that the global replication timing program is lost in *rif1*Δ (Fig. 4.12). Similarly to the RT changes at subtelomeres, the change in the global program cannot be explained by a change in the global distribution of origin firing efficiencies of *rif1*Δ origins (Fig. 4.6). This implies that, unlike the local changes seen at Rif1 BSs, the global RT program is not caused exclusively by changes in origin firing. Given that the global RT program depends on Rif1, it is possible that Rif1 may have another role in controlling *S. pombe* replication timing, which is independent from its role in origin inhibition. We have, therefore, decided to re-focus our analyses on the factors influencing the global RT program of *S. pombe* and the role Rif1 plays in this pathway.

Chapter 5

5.1 Introduction

Replication timing (RT) profiles across a number of metazoan backgrounds revealed that their genomes are organised into replication timing domains, whose sizes range from hundreds of kilobases to several megabases (Rhind and Gilbert 2013). On a population level, the RT across replication timing domains is constant (Rivera-Mulia and Gilbert 2016). The temporal order in which these replication domains are replicated is referred to as the global RT program.

Similarly to metazoans and other lower eukaryotes (Cornacchia, Dileep et al. 2012, Hiraga, Alvino et al. 2014), Rif1 has been implicated as a regulator of global RT program in *S. pombe* (Hayano, Kanoh et al. 2012, Kanoh, Matsumoto et al. 2015). While in metazoans Rif1 has been suggested to underpin the nuclear architecture responsible for the global RT program (Foti, Gnan et al. 2015) (discussed in 1.4.2), no such role of Rif1 has been described in yeast.

5.1.1 Action of Rif1 in yeast

Rif1 in *S. pombe* has been implicated in a number of seemingly disparate cellular processes. Similarly to its *S. cerevisiae* orthologue (Wotton and Shore 1997), Rif1 is recruited to telomeres (Kanoh and Ishikawa 2001). In *S. pombe*, however, this recruitment is mediated by Taz1 (Kanoh and Ishikawa 2001) and Rif1 activity is not involved in preventing aberrant chromosome fusion by telomere capping (Miller, Ferreira et al. 2005). In *S. pombe*, *rif1*Δ results in only a mild telomere elongation phenotype (Kanoh and Ishikawa 2001). As discussed in 4.1.2, Rif1 in *S. pombe* has been shown to contribute to the formation of heterochromatin at facultative heterochromatin islands (Zofall, Smith et al. 2016). Loss of Rif1 has been reported to affect spore formation (Kanoh and Ishikawa 2001) and this is likely linked to the disruption of the non-DSR islands associated with meiotic genes. Unlike in higher eukaryotes, where Rif1 is a crucial effector of 53BP1 promoting non-homologous end joining (Zimmermann and de Lange 2014), *rif1*Δ in *S. pombe* was shown to not increase sensitivity when cells were exposed to a range of DNA damaging agents (Hayano, Kanoh et al. 2012), suggesting it is may not be involved in the DNA damage response.

Rif1 was also described as a regulator of global origin firing in *S. pombe*. Its absence led to the apparent suppression of early firing origins and stimulation of late and / or dormant origins (Hayano, Kanoh et al. 2012). The abrogation of Rif1 binding to a single site, was reported to affect the origin firing as far as 50 kb away (Kanoh, Matsumoto et al. 2015).

It was also noted that *rif1*Δ was an efficient bypass mutant to *hsk1*Δ lethality (Hayano, Kanoh et al. 2012). In *S. cerevisiae*, a similar phenotype was noted and was shown to be dependent on the Rif1 mediated recruitment of the PP1 (protein phosphatase 1) Glc7 (Hiraga, Alvino et al. 2014). Rif1 recruits PP1 phosphatases via its conserved SILK/RVxF domains. The interaction between the phosphatases and Rif1 alleles that contain point mutations across the SILK/RVxF domains (Rif1-PP1) are disrupted in *S. cerevisiae* and *S. pombe* (Dave, Cooley et al. 2014). *S. pombe* contains two PP1 phosphatases - Dis2 and Sds21, whose simultaneous deletion is lethal (Alvarez-Tabares, Grallert et al. 2007)

In *S. cerevisiae*, the Rif1-Glc7 interaction was shown to lead to a Glc7 dependent decrease in Mcm2-7 phosphorylation (Hiraga, Alvino et al. 2014). In their model, (Hiraga, Alvino et al. 2014) propose that this de-phosphorylation leads to the inhibition of origin firing in G_1. They suggest that the inhibition is lifted in S-phase by the abrogated interactions between Rif1 and Glc7, caused by the phosphorylation of the Rif1 N-terminus by DDK. This model suggests that a non-phosphorylatable allele of Rif1 would exhibit stronger interactions with the phosphatases than wild type Rif1. This effect, however, was not observed in *S. pombe* (Dave, Cooley et al. 2014). An allele of Rif1 with mutations across seven predicted sites of CDK and DDK phosphorylation (Rif1-7A) did not interact more strongly with PP1 phosphatase Sds21 than wild type Rif1 (Dave, Cooley et al. 2014).

The ability of Rif1 to affect origin firing by the indirect dephosphorylation of Mcm2-7, has been suggested as a mechanism for controlling the global RT program (Kanoh, Matsumoto et al. 2015). This hypothesis, however, is based on the assumption that modulating origin firing efficiencies is responsible for the global RT pattern.

5.1.2 Application of current mathematical models of replication to *S. pombe*

As discussed in 1.4.1, the currently accepted mathematical model (Yang, Rhind et al. 2010) explaining global RT in *S. cerevisiae* is considered to be descriptive of global RT in higher eukaryotes (Rhind and Gilbert 2013). The model is built on two assumptions. Firstly, origin efficiencies were defined based on firing-time distributions, i.e., the firing of efficiently and inefficiently firing origins was restricted to early and late S-phase, respectively. Secondly, the efficiency of all unreplicated origins increased towards the end of S-phase (Yang, Rhind et al. 2010). This model was used to simulate DNA replication across all 16 *S. cerevisiae* chromosomes *in silico* and generated accurate RT profiles for all of them (Yang, Rhind et al. 2010). This reconciled the stochasticity of origin firing with the constant replication timing profiles (Bechhoefer and Rhind 2012).

It is important to note, that although the RT profiles generated were global, they only spanned the relatively short *S. cerevisiae* chromosomes (whose largest chromosome is 1.5 Mb). In Chapter 4, we showed that a local increase in the origin activity around Rif1 BSs and Taz1 dependent heterochromatin islands increased locally (Figs. 4.8 A and 4.5 A , respectively) affected the relative RT over 10s-100s of kbs around the sites (Fig. 4.13), which is in accordance with the (Yang, Rhind et al. 2010) model. We also noted, however, that the increase the relative RT of subtelomeres in *taz1*Δ (Fig. 4.10 A) and *rif1*Δ (Fig. 4.12), as well as the loss of the entire global RT program in *rif1*Δ (Fig. 4.12), were not correlated with concomitant changes in origin firing activity around subtelomeres (Fig. 4.7) or the global distribution of origin firing efficiencies (Fig. 4.1). While linking the absolute origin firing efficiencies to the temporal order in which the origins fire results in accurate RT profiles over relatively short distances (several megabases) (Yang, Rhind et al. 2010), this may not be the case over larger chromosomes or globally.

Based on our data, we carried out an analysis to determine whether the global replication timing program and origin firing efficiencies are linked in *S. pombe*. Similarly to *S. cerevisiae* (Czajkowsky, Liu et al. 2008), *S. pombe* origins also fire stochastically (Patel, Arcangioli et al. 2006). The chromosomes are, however, much longer, with the shortest chromosome being 3.5 Mb (Wood, Gwilliam et al. 2002)

and the replication timing domains spanning over 2 Mb (Fig. 4.9), making them more similar to those in higher eukaryotes.

5.1.3 Current methods to determine global RT in yeast

The Pu-Seq RT profiles shown in Chapter 4 are an indirect measure of RT and are calculated from data that describe the progression of leftward moving replication forks (calculations described in 2.2.4.3). To study the global RT program further we wanted to generate RT profiles independent of Pu-Seq data.

A number of deep-sequencing methods have been developed to explore the global RT program in *S. cerevisiae* (Muller, Hawkins et al. 2014). All of these methods are based on comparing whole genome sequencing data between exponentially growing cells in S-phase and stationary cells, which are mostly in G_2. Several different ways of obtaining synchronous cultures, some of which do not rely on checkpoint mutants, have been compared and the resulting RT profiles were very similar (Muller, Hawkins et al. 2014).

To generate Sort-Seq RT profiles, fixed cells from exponentially growing cultures are divided into S-phase and G_2 populations using fluorescence-activated cell sorting (FACS) (Muller, Hawkins et al. 2014). Cell sorting based on DNA content is straightforward in *S. cerevisiae*, where DNA content per cell increases with the progression of the cell cycle, i.e., G_2 cells contain twice the amount of DNA compared to G_1 cells. In *S. pombe*, however, cytokinesis does not occur until after G_1. This results in G_1 cells with the same DNA content as G_2 cells (a schematic of the DNA content changes in the *S. pombe* cell cycle is shown in Fig. 5.1). A more complicated method to analyse the *S. pombe* cell cycle using FACS has been developed (Knutsen, Rein et al. 2011) and successfully applied to produce Sort-Seq RT profiles for *S. pombe* (Fig. 4.9).

Cell collection for marker frequency analysis (MFA) is simpler. MFA RT profiles are calculated by comparing sequencing data from asynchronous exponentially growing cells to that from cells in stationary cultures. Although MFA data are noisier, smoothed MFA RT profiles still reveal the same global RT profiles as Sort-Seq (Muller, Hawkins et al. 2014).

Accuracy of MFA can be increased by using a partially synchronised exponentially growing population. *S. cerevisiae* cultures are routinely synchronised using the mating pheromone α-factor, which is naturally secreted by cells to block cell division in G_1 and induce expression of mating-specific genes (Breeden 1997). When G_1 cells are released from the α-factor block, they progress synchronously through the cell cycle. Their progression can be tracked using FACS to determine the time points at which the culture was in S-phase. In this version of MFA, gDNA from all S-phase time points is extracted and sequenced. All of the S-phase data sets are then separately compared to the sequencing data generated from the G_1 arrested cells, resulting in an RT profile for each of the time points. These RT profiles are then analysed together to produce a median replication time profile, which is more accurate than the "noisy" MFA profile and compares well with Sort-Seq data (Muller, Hawkins et al. 2014).

A modified version of MFA was used to produce an accurate global RT profile of *S. pombe* (Daigaku, Keszthelyi et al. 2015). Instead of α-factor synchronisation, which does not work in *S. pombe*, cells were synchronised using elutriation. Centrifugal elutriation separates the cells based on size, allowing the collection of large amounts of small G_2 cells (Hagan, Grallert et al. 2016). Following elutriation, the G_2 cells were released into a synchronous cell cycle, whose progression was followed using FACS and microscopy. gDNA from S-phase time points was subsequently extracted, sequenced and analysed following the same methods as those described for *S. cerevisiae* by (Muller, Hawkins et al. 2014).

Although both Sort-Seq and "accurate" MFA have been modified and used successfully in *S. pombe*, the protocols involved are both expensive and time consuming.

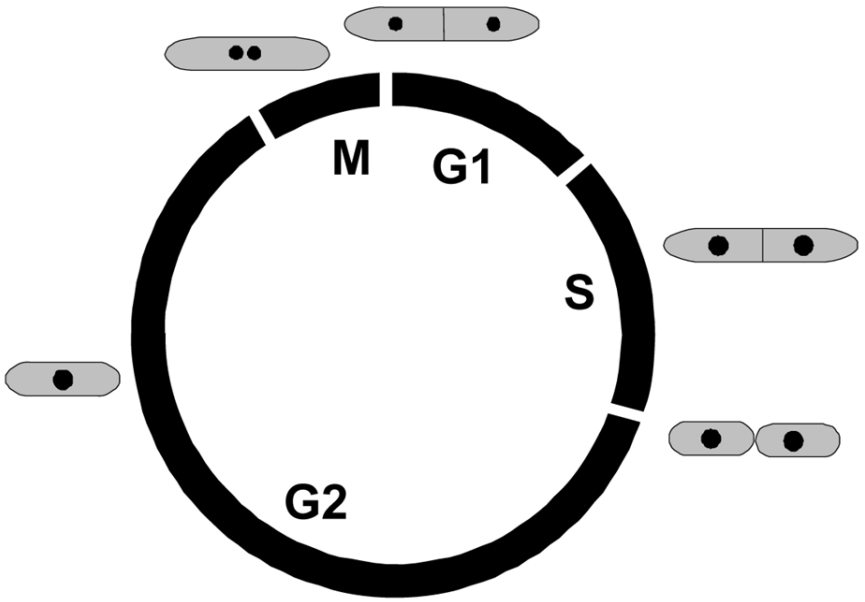


Fig. 5.1 - ***S. pombe*** **morphology and DNA content at each stage of the cell cycle**

The small G_2 cells contain a single 2N nucleus. The cells then undergo mitosis, before cytokinesis, resulting in M-phase cells with two 1N nuclei. Cells in G_1 and S-phase both have a septum and two nuclei (1N in G_1 and 2N after S-phase). G_2 cells directly post-division are still attached but without a visible septum between the two small cells. Adapted from (Knutsen, Rein et al. 2011).

5.1.4 Chapter aims

To validate the phenotype of the loss of global RT program in *rif1*Δ, we developed a simplified and more cost effective deep-sequencing technique (Elutri-Seq) to generate RT profiles in *S. pombe*.

We generated and analysed *rif1*Δ data to recapitulate the loss of global RT phenotype using Elutri-Seq. Simultaneously, we also generated a wild type Elutri-Seq RT profile using previously published sequencing data, as a proof of concept.

Next, we explored the effect of the Rif1 mediated PP1 phosphatase recruitment and Rif1 phosphorylation on the global RT and origin usage by carrying out Pu-Seq and Elutri-Seq analyses on the previously described Rif1-PP1 and Rif1-7A alleles (Dave, Cooley et al. 2014).

Finally, to determine whether global RT is driven by the efficiency of origin firing, we analysed the differences between origin usage in early and late replicating regions in *S. pombe*.

5.2 Results

In order to validate the loss of global RT program seen in *rif1*Δ using Pu-Seq (Fig. 4.12), we developed Elutri-Seq, a composite of the Sort-Seq and MFA methods, to generate global RT profiles in *S. pombe*.

5.2.1 Elutri-Seq RT profiles for *rif1*Δ and wild type *S. pombe*

As was done for *S. pombe* MFA (Daigaku, Keszthelyi et al. 2015), *rif1*Δ cells were synchronised in G_2 using centrifugal elutriation. $2.5x10^7$ G_2 cells were harvested and re-suspended in rich YE media to a final concentration of $2.5x10^5$ cells/mL. The synchronisation 20 minutes after elutriation is shown in Fig. 5.2 A. The synchrony with which the cells progressed through the cell cycle was measured by the counting percentage of cells in each phase at every time point (Fig. 5.2 B) and using FACS (Fig. 5.2 C).

The cell cycle progression of *rif1*Δ cells (Fig. 5.2) was similar to that of wild type (wild type data reported by (Daigaku, Keszthelyi et al. 2015)). The peak of septation (marker of early S-phase) was seen earlier in *rif1*Δ - 60 minutes after release (Figs. 5.2 B and C), compared to 75 minutes for wild type (Daigaku, Keszthelyi et al. 2015).

Unlike for MFA, we did not analyse all of the S-phase time points. Instead, we took the Sort-Seq approach and compared a single S-phase population with the G_2 population. gDNA from the 60 minute (S-phase) and 20 minute (G_2) time points was extracted, sonicated and used to prepare Illumina libraries (optimised protocol in 2.2.5.1). The libraries were paired-end sequenced and the reads generated were aligned to the SP2 reference genome. The resultant coverage of the *S. pombe* genome was calculated using the Lander/Waterman equation (Lander and Waterman 1988) and is shown in Table 5.1. The coverage of the *S. cerevisae* genome used for Sort-Seq (Muller, Hawkins et al. 2014), was also calculated and is shown for comparison in Table 5.1.

Table 5.1 – **Number of reads mapped and total coverage for *S. pombe rif1Δ* Elutri-Seq and wild type *S. cerevisiae* Sort-Seq** (the latter carried out by (Muller, Hawkins et al. 2014)).

The total number of mapped reads is the sum of R1 and R2 mates that aligned in pairs (concordantly and discordantly) and single reads that aligned in single end mode.

Experiment and organism	Size of haploid genome/Mb	Length of reads/bp	Sample	Total number of mapped reads	Coverage
Elutri-Seq (*S. pombe*)	12.57	81	S-phase	139,287,892	897.6
			G_2	119,210,264	768.2
Sort-Seq (*S. cerevisiae*)	12.1	100	S-phase	155,734,122	1287.1
			G_2	131,816,351	1089.4

The Elutri-Seq coverage was lower than, but still comparable, to that of *S. cerevisiae* Sort-Seq. We continued the analysis without increasing the depth of sequencing.

To validate the Elutri-Seq approach, we generated a wild type Elutri-Seq RT profile in parallel. The wild type profile was generated by analysing two time points from the *S. pombe* MFA data (Daigaku, Keszthelyi et al. 2015). The 75 and 125 minute time points were chosen as S-phase and G_2 samples, respectively. Both were chosen to best match the DNA content and number of S-phase/G_2 cells seen in the

*rif1*Δ S-phase and G_2 samples. The reads from the wild type S-phase and G_2 libraries were mapped to the SP2 reference genome and coverage was calculated using the Lander/Waterman equation (Lander and Waterman 1988) (Table 5.2).

Table 5.2 – **Number of reads mapped and total coverage for *S. pombe* MFA** done by (Daigaku, Keszthelyi et al. 2015)

The total number of mapped reads is the sum of R1 and R2 mates that aligned in pairs (concordantly and discordantly) and single reads that aligned in single end mode.

Experiment and organism	Size of haploid genome/Mb	Length of reads/bp	Sample	Total number of mapped reads	Coverage
MFA (*S. pombe*)	12.57	81	S-phase (75min)	107,000,844	689.5
			G_2 (125min)	102,287,120	659.1

Sort-Seq RT profiles are generated by calculating the ratio between the number of reads that map to each locus (i.e., the copy number) in early S-phase and G_2 (Muller, Hawkins et al. 2014). In the replicating sample, the copy number of each locus is determined by when in S-phase the locus replicates. Loci that replicated earlier will generate more reads and therefore result in a greater copy number than those that replicate later in S-phase. Fully replicated G_2 cells have an equal copy number at each locus.

We generated Elutri-Seq profiles following the same logic, using standard samtools commands (shown and described in 2.2.5.2). Briefly, the SAM files containing the mapped reads in the S-phase and G_2 datasets were converted into BAM format, sorted and binned into windows of 1 kb. The 5' end of each read was assigned to a bin and the total number of reads in each bin was counted. The counts in each bin were then normalised to the total number of counts in each sample. A ratio between the normalised S-phase and G_2 counts in each bin was calculated to produce the final Elutri-Seq Trep. Density distributions of the Elutri-Seq Trep are shown in Fig. 5.3.

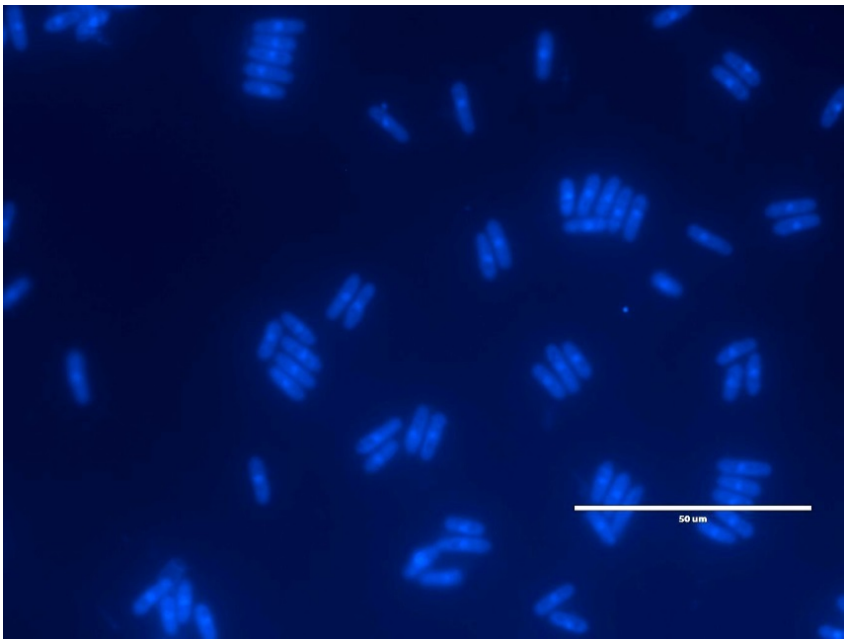
The distribution of wild type Elutri-Seq Trep was skewed towards late replication, with a small population of early replicating loci forming a "bulge" that was skewed towards early replication. This is descriptive of wild type *S. pombe* replication where much of the genome is replicated late, with only the smallest chromosome and parts of Chromosomes 1 and 2 replicating early (Fig. 4.9). *rif1*Δ Elutri-Seq Trep, on the other hand, was normally distributed, suggesting that in *rif1*Δ all loci had an equal probability of being replicated during early S-phase.

To compare the distributions directly, the wild type and *rif1*Δ Elutri-Seq Trep were normalised to fraction of the genome replicated in the S-phase sample used (method described in 2.2.5.3). The normalised Elutri-Seq Trep was plotted for each bin (Fig. 5.4 A), to determine which areas of the genome contributed to early replicating "bulge" on the wild type Trep distribution shown in Fig. 5.3. Superimposed density distributions of normalised wild type and *rif1*Δ Elutri-Seq Trep are shown in Fig. 5.4 B.

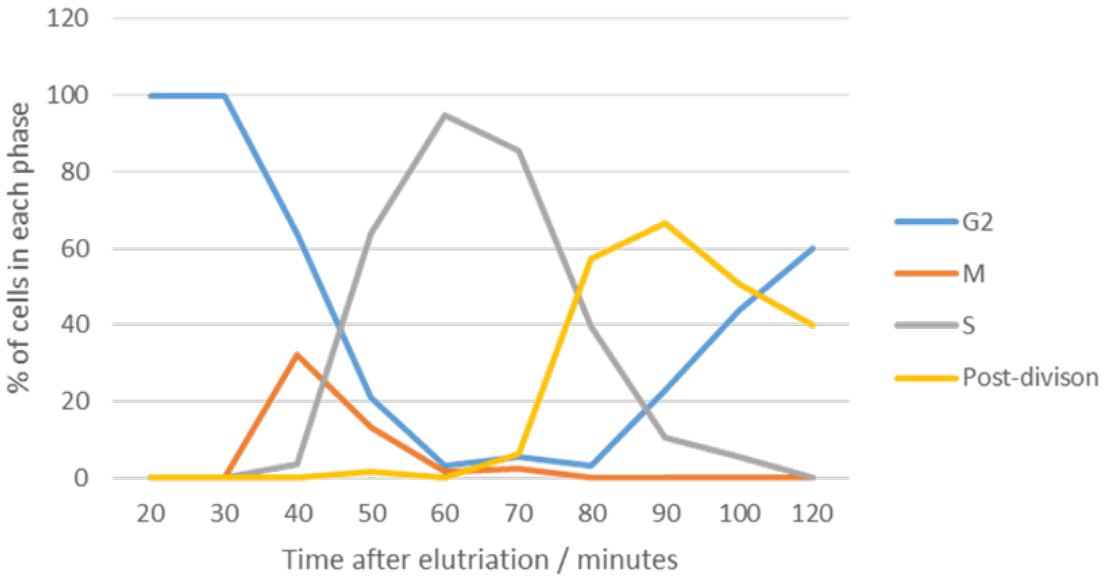
The wild type Elutri-Seq Trep (Fig. 5.4 A) recapitulated the wild type Sort-Seq data (Fig. 4.9). Chromosome 3 and the same parts of Chromosomes 1 and 2 were early replicating (Fig. 5.4 A). This global RT program was lost in *rif1*Δ. Similarly to the *rif1*Δ Pu-Seq RT profile (Fig. 4.12), the Elutri-Seq *rif1*Δ RT profile was flat (Fig. 5.4 A). This showed that in *rif1*Δ, the hierarchy of replication timing is abolished and no parts of the genome replicate earlier than others.

The good correlation between wild type Elutri-Seq and Sort-Seq RT profiles, shows that Elutri-Seq is a good substitute to the more expensive and time consuming methods described in 5.1.3 for analysing RT in *S. pombe*. The *rif1*Δ Elutri-Seq data validated the Pu-Seq Trep, which also showed a loss of the global RT program.

A) Synchronisation after 20 minutes



B) Septation index



C) FACS

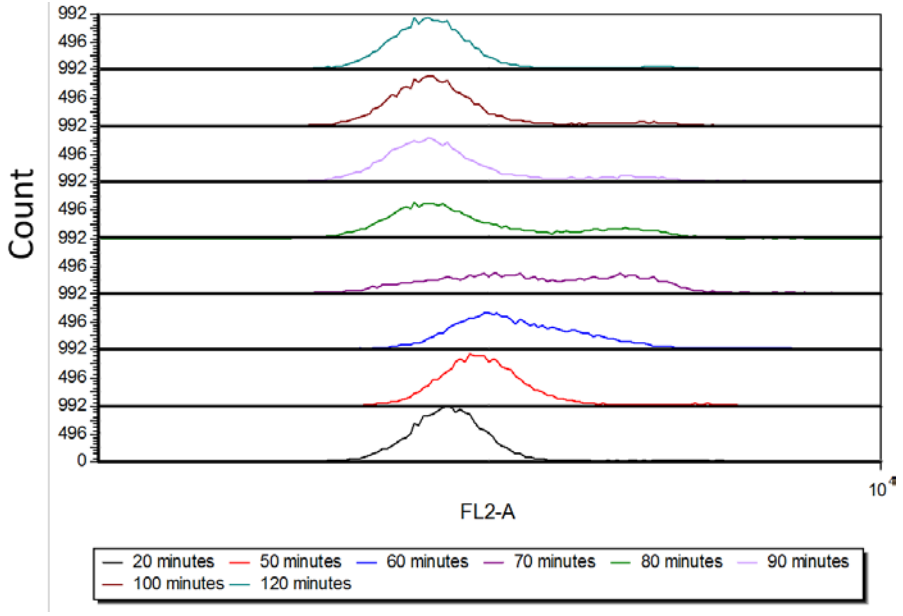


Fig. 5.2 - **Analysis of the cell cycle of *rif1*Δ cells after elutriation**

A) Synchronisation of *rif1*Δ 20 minutes after elutriation. Cells were stained with 1 µg/mL 4',6'-diamidino-2-phenylindole (DAPI) and 2.5% v/v calcofluor-white (which stain nucleic acid and the septum, respectively). Cells were visualized using an inverted fluorescence microscope (EVOS™ FL).

B) The synchronous passage of *rif1*Δ cells through the cell cycle after elutriation. Cells were stained and visualised at each time point (as described in A). Cells were counted and divided into different phases of the cell cycle based on their morphology - cells with one nucleus - G_2 ; cells with two nuclei - M-phase ; cells with a septum- S-phase ; two cells joined, without a visible septum, and with one nucleus each - post-division.

C) FACS analysis of cells at each time point. 1.25×10^6 cells were collected at each time point, stained with propidium iodide and analysed for DNA content on BD Accuri™ C6 Plus flow cytometer.

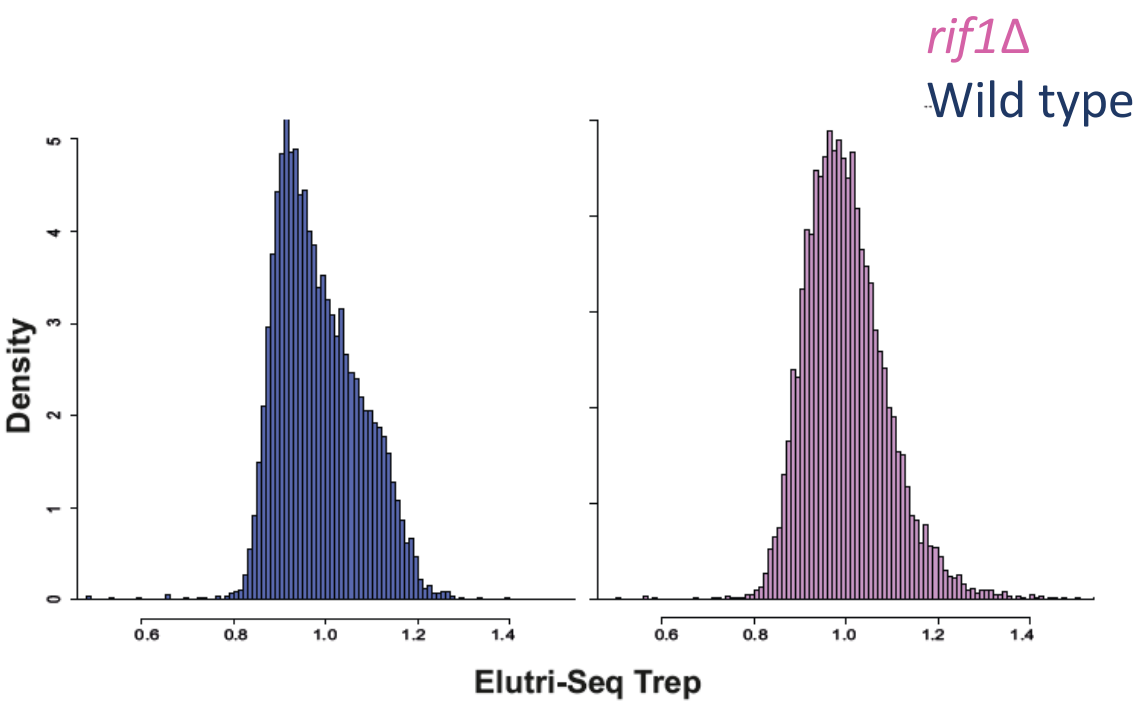


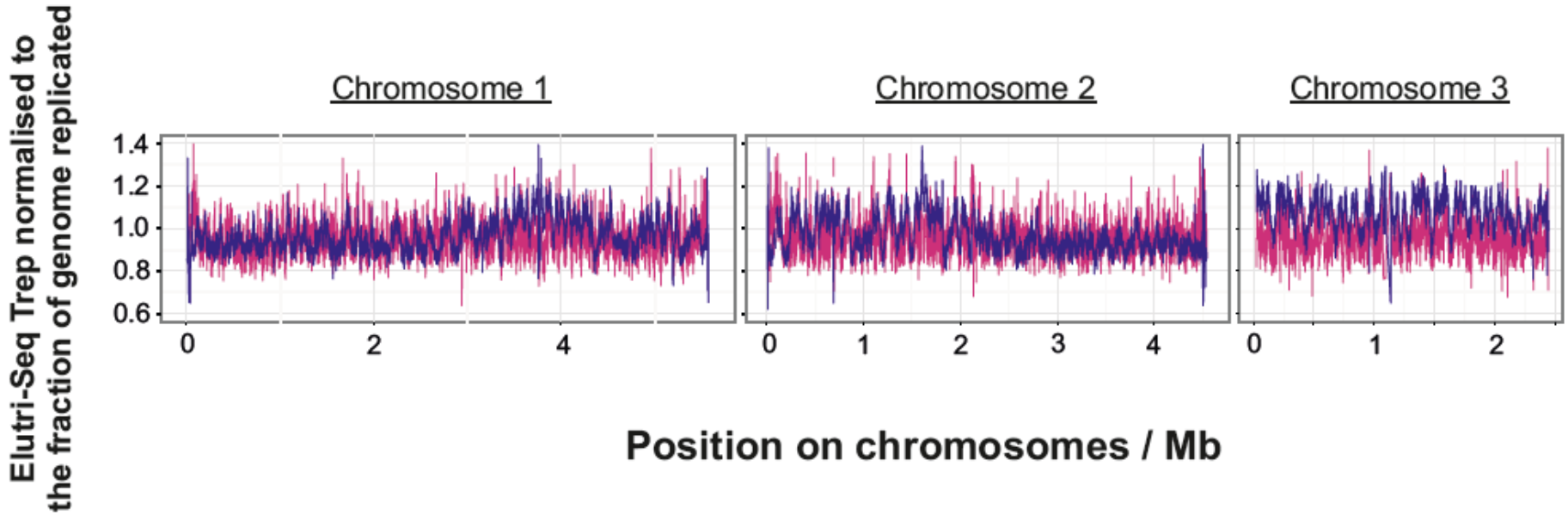
Fig. 5.3 - **Density distribution of Elutri-Seq Trep for wildtype and *rif1*Δ *S. pombe***

Reads from S-phase and G_2 libraries (sequenced in both backgrounds for a coverage over 650 X) were aligned to the SP2 reference genome, binned into 1 kb windows, the number of 5' ends of reads was counted in each bin and normalised to the total number of counts. The ratio the S-phase and G_2 counts in every bin was calculated to produce Elutri-Seq Trep. The x-axis represents the relative replication time of each locus in the genome, where higher ratio values represent earlier replicating loci.

Wild type Elutri-Seq Trep (blue) is not normally distributed - large regions of the genome replicate earlier than others. *rif1*Δ Elutri-Seq Trep (pink) is normally distributed, showing that all regions have an equal probability of being replicated at any point during S-phase in *rif1*Δ. The *rif1*Δ and wildtype histograms were binned into 1000 and 500 bins, respectively.

A) Elutri-Seq genome wide

rif1Δ
Wild type



B) Distribution of Elutri-Seq

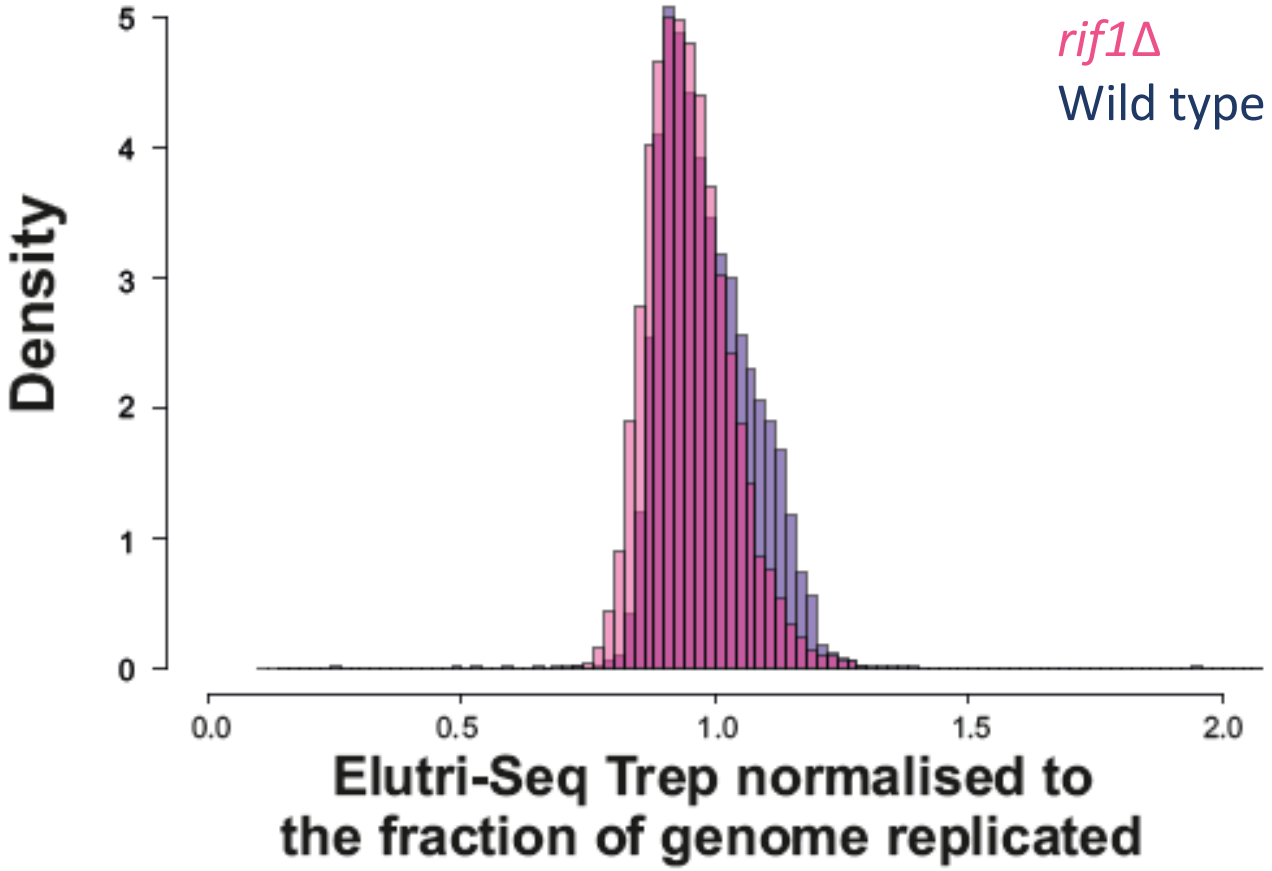


Fig. 5.4 - **Wild type and *rif1*Δ Elutri-Seq Trep, normalised to the fraction of the genome replicated**

The wild type and *rif1*Δ Elutri-Seq RT profiles were normalised to the fraction of the genome replicated (as described in 2.2.5.3).

A) The Elutri-Seq Trep, normalised to the fraction of the genome replicated, plotted against each bin.

B) Superimposed density distributions of wild type and *rif1*Δ Elutri-Seq Trep (normalised to the fraction of the genome replicated). The *rif1*Δ and wildtype histograms were binned into 500 and 250 bins, respectively.

5.2.2 The effect of different Rif1 alleles on replication timing and origin firing

Rif1 has been shown to interact with PP1 phosphatases via its SILK/RVxF domains (Dave, Cooley et al. 2014, Hiraga, Alvino et al. 2014). In *S. cerevisiae*, the phosphorylation of Rif1 inhibits the interaction of Rif1 with the phosphatases. (Hiraga, Alvino et al. 2014). Rif1-PP1 and Rif1-7A are *S. pombe* Rif1 alleles that are mutated in the SILK/RVxF domains and at putative Rif1 phosphorylation sites, respectively (Cooley, Dave et al. 2014). A schematic showing the hypothesised effects of each allele is shown in Fig. 5.5 A. To further explore the role of Rif1 in *S. pombe*, we went on to test the effect of these Rif1 alleles on the local inhibitory origin firing inhibition around Rif1 BSs and the maintenance of the global RT program.

5.2.2.1 Effect of Rif1-PP1 and Rif1-7A on origin firing and local RT

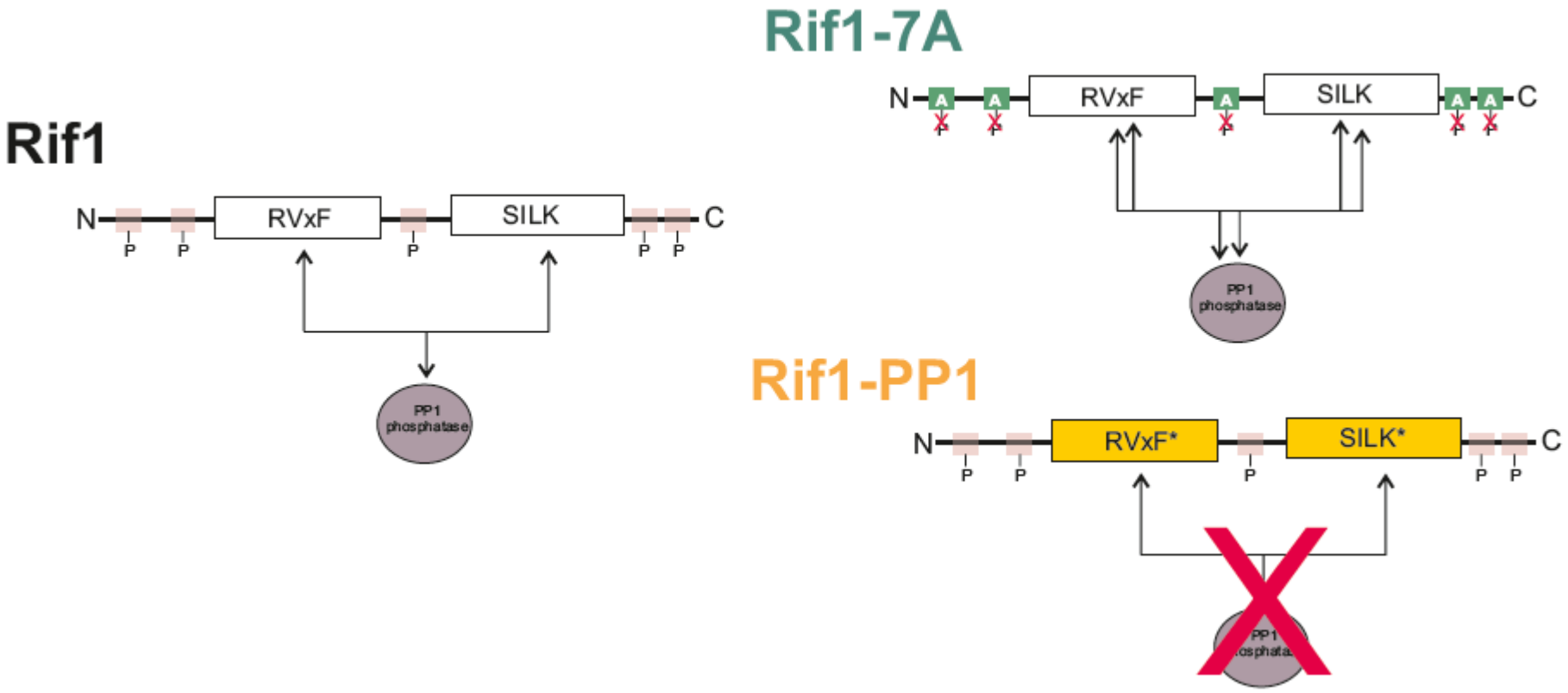
The distribution of origin firing efficiencies in both Rif1-7A and Rif1-PP1 remained bimodal (Fig. 5.5 B). The numbers of origins detected were within the wild type range (1075 and 994 origins of replication for Rif1-7A and Rif1-PP1, respectively; Rif1-PP1 data are an average of two independent biological repeats).

Next, we analysed the origin activity up and downstream of strong Rif1 BSs and Taz1 dependent heterochromatin islands in Rif1-PP1 and Rif1-7A. Origin activity around both sets of sites increased in Rif1-PP1 (Figs. 5.6 A and B). The origin activity around Rif1 BSs was, however, significantly lower in Rif1-PP1 compared to *rif1*Δ. The Rif1-PP1 allele has been previously described as incompletely penetrant (with regards to telomere elongation) (Zaaijer, Shaikh et al. 2016). It is, therefore, possible that the difference in the origin activation between the Rif1-PP1 and *rif1*Δ is due to residual PP1 interactions. The increase in origin activity resulted in an increase in local RT around strong Rif1 BSs and Taz1 dependent heterochromatin islands in Rif1-PP1 (Figs. 5.6 C and D). These data show that the interaction of Rif1 with phosphatases is necessary to inhibit origin firing, supporting the model of Mcm2-7 de-phosphorylation (Hiraga, Alvino et al. 2014).

We saw no increase in origin activity (or related changes in local RT) in Rif1-7A (Fig. 5.6). It is important to reiterate that the Rif1-7A allele did not show a significantly stronger interaction with the phosphatases, compared to wild type

(Dave, Cooley et al. 2014) and may, therefore, not be an accurate model for a complete loss of phosphorylation on Rif1.

A) Rif1 alleles



B) Origin efficiency in Rif1-PP1 and Rif1-7A

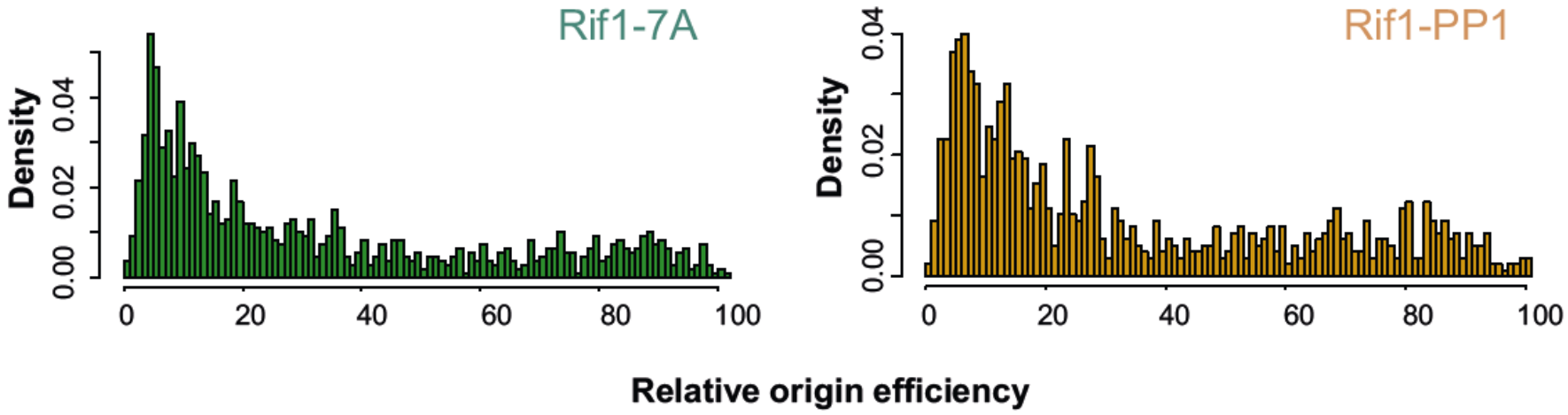
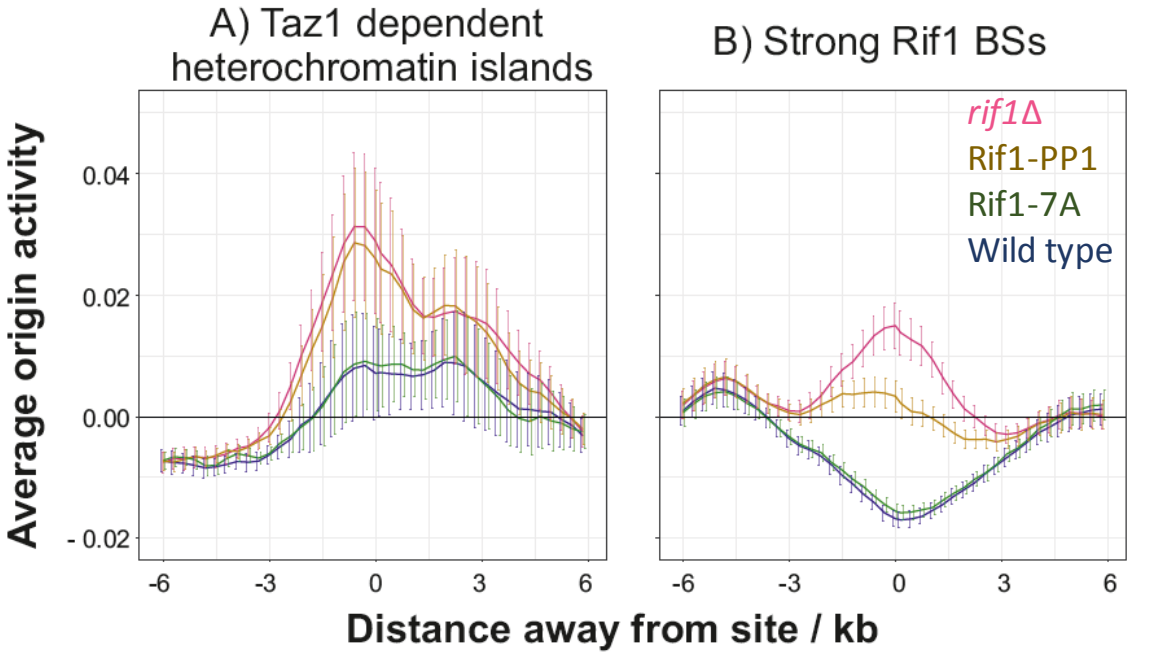


Fig. 5.5 - **Simplified interactions of the Rif1 alleles with PP1 phosphatases and the density distribution of origin firing efficiencies in Rif1-PP1 and Rif1-7A**

A) Schematic of Rif1 alleles. Wild type Rif1 interacts with phosphatases via its RVxF and SILK domains. The Rif1-PP1 allele contains four point mutations (two in each domain) that convert four residues to alanine. The resultant Rif1 protein cannot interact with PP1 phosphatases Dis2 and Sds21. The phosphorylation of Rif1 by CDK and DDK in S-phase has been proposed to inhibit the interaction with the phosphatases in *S. cerevisiae* (Hiraga, Alvino et al. 2014), which would stop the de-phosphorylation of Mcm2-7 and allow origins to fire. The mutation of seven putative CDK and DDK phosphorylation sites (Rif1-7A) was proposed to result in a Rif1 whose interactions with the phosphatases were enhanced.

B) Density distribution of origin firing efficiencies in Rif1-PP1 and Rif1-7A *S. pombe*. Origins of replication were mapped for each strains using Pu-Seq, applying a threshold of 0.3, i.e., positive peaks of the differentials of polymerase usage whose heights were above the 30th percentile were mapped as origins. The efficiency of firing was normalised to the value of the 99th percentile in each data set.


A) Taz1 dependent heterochromatin islands
B) Strong Rif1 BSs
rif1Δ
Rif1-PP1
Rif1-7A
Wild type
Average origin activity
0.04
0.02
0.00
- 0.02
-6
-3
0
3
6
Distance away from site / kb

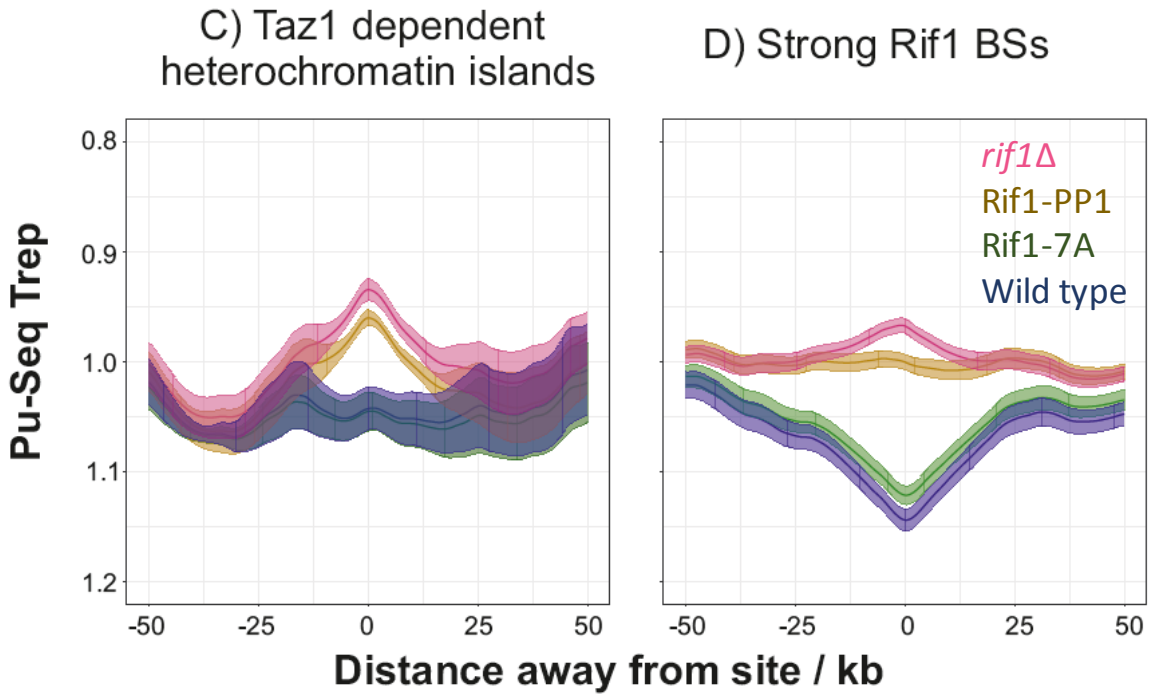

C) Taz1 dependent heterochromatin islands
D) Strong Rif1 BSs
rif1Δ
Rif1-PP1
Rif1-7A
Wild type
Pu-Seq Trep
0.8
0.9
1.0
1.1
1.2
-50
-25
0
25
50
Distance away from site / kb

Fig. 5.6 – **The average origin activity and Pu-Seq Trep around Taz1 dependent facultative heterochromatin islands and strong Rif1 BSs in *rif1*Δ, Rif1-7A, Rif1-PP1 and wild type backgrounds**

The mean origin activity (i.e., the average of the differentials of the polymerase δ usage on the reverse strand and polymerase ε forward strand) in each 300 bp bin is shown for 6 kb up and downstream of the midpoint of each site. The error bars represent the standard error of the mean in each bin. The origin activity data shown here were calculated from two independent biological repeats for all backgrounds except Rif1-7A.

Pu-Seq Trep was calculated using the progression of leftward moving forks and assuming a constant fork velocity of 1.5 kb/min. It represents when in S-phase each locus is replicated. The y-axis was, therefore inverted to maintain the convention of early and late replicating regions being shown on the top and bottom of the graph, respectively. Pu-Seq Trep 6 kb up and downstream of the midpoint of each site was recorded. The Pu-Seq Trep data shown here were calculated from two independent biological repeats for all backgrounds except Rif1-7A. The error bars represent the standard error of the mean in each bin.

A) The mean activity in each bin averaged across the 6 Taz1 dependent heterochromatin islands. The wild type and *rif1*Δ origin activity data shown here are identical to those shown in Fig. 4.5 A and are only shown here for comparison.

B) The mean activity in each bin averaged across the 35 strong Rif1 binding sites. The wild type and *rif1*Δ origin activity data shown here are identical to those shown in Fig. 4.8 A and are only shown here for comparison.

C) The Pu-Seq Trep in each bin averaged across the 6 Taz1 dependent heterochromatin islands. The wild type and *rif1*Δ origin activity data shown here are identical to those shown in Fig. 4.13 A and are only shown here for comparison.

D) The Pu-Seq Trep in each bin averaged across the 35 strong Rif1 binding sites. The wild type and *rif1*Δ origin activity data shown here are identical to those shown in Fig. 4.13 B and are only shown here for comparison.

5.2.2.2 Effect of Rif1-PP1 and Rif1-7A on global RT

The Rif1-PP1 Pu-Seq RT Trep profile showed a loss of global RT (Fig. 5.7), identical to that seen in *rif1*Δ (Fig. 4.12), showing that the maintenance of the global RT program is dependent on the interaction of Rif1 with the PP1 phosphatases. The Rif1-7A Pu-Seq Trep RT profile was wild type (Fig. 5.7), suggesting that the phosphorylation of Rif1 did not affect global RT. The Pu-Seq Trep RT profiles shown in Fig. 5.7 show the RT across Chromosome 2, where the loss of wild type global RT program is most visible (early and late replicating domains disappear).

To validate the loss of the global RT program in Rif1-PP1, we carried out Rif1-PP1 Elutri-Seq, as described for *rif1*Δ in 5.2.1. Cells were synchronised by elutriation and the collected G_2 cells were resuspended to a final volume of $2.5x10^5$ cells/mL. The synchronisation 20 minutes after elutriation is shown in Fig. 5.8 A. The synchrony with which the cells progressed through the cell cycle was measured by the counting percentage of cells in each phase at every time point (Fig. 5.8 B) and using FACS (Fig. 5.8 C).

The synchronisation of Rif1-PP1 was successful, but not to the same extend as for *rif1*Δ. 20 minutes after Rif1-PP1 elutriation, over 97% of cells were in G_2 (Fig 5.8 B). While this level of synchronisation is much better than could be achieved using other *S. pombe* methods (such as lactose gradients or even checkpoint mutants), it was lower than the 100% synchrony achieved for *rif1*Δ (Fig. 5.2 B). This difference resulted in over 16% of cells being in G_2 at the peak of septation (70 minutes after elutriation). To avoid confounding the Elutri-Seq S-phase data, we chose the 60 minute sample instead. Although there were fewer septated cells 60 minutes after elutriation, the proportion of G_2 cells was also lower (under 10%). Given the lack of synchrony at the 20 minute time point, the 120 minute time point was used as the Rif1-PP1 G_2 sample for Elutri-Seq.

gDNA from the two samples was extracted, sonicated and used to prepare Illumina libraries (as described in 2.2.5.1). Unfortunately, the alignment rate and, therefore, the final coverage were very low (Table 5.3). Given that the Elutri-Seq profiles are a ratio and the counts in each bin are normalised to the total number of counts per sample, we assumed a lower coverage would not significantly affect the results.

Table 5.3 – **Number of reads mapped and total coverage for Rif1-PP1 Elutri-Seq**

The total number of mapped reads is the sum of R1 and R2 mates that aligned in pairs (concordantly and discordantly) and single reads that aligned in single end mode.

Length of reads/bp	Sample	Total number of reads	Overall alignment rate	Number of reads mapped to the reference genome	Coverage
80	S-phase	124,988,114	0.84%	1,049,900	6.8
	G_2	101,143,955	2.86%	2,892,717	18.6

The obtained reads were analysed and normalised as described in 5.2.1. The Elutri-Seq Trep ratio was normally distributed (Fig. 5.8 D), similarly to that in *rif1*Δ (Fig. 5.3), validating the loss of the global RT program shown by Pu-Seq Trep (Fig. 5.7). Due to the substantially lower depth of sequencing, however, plotting the Eluri-Seq Trep in each bin did not provide any meaningful information (data not shown). From this we concluded that low coverage Elutri-Seq data can be used to analyse large changes to global RT program. The resolution, however, is not sufficient to be able to determine the exact genomic loci that contribute to the changes.

Taken together, the data show that the loss of Rif1 interactions with PP1 phosphatases increases origin activity locally (Figs. 5.6 A and B), which leads to regional change in RT (Figs. 5.6 C and D). Additionally, this interaction is also important to maintain the global RT program (Figs. 5.7 and 5.8 D). As was seen in *rif1*Δ (Fig. 4.1), the Rif-PP1 distribution of origin firing efficiencies was wild type (Fig. 5.5 B). These data further reinforce the disparity between the local effects of modulating origin firing and the maintenance of the global RT program.

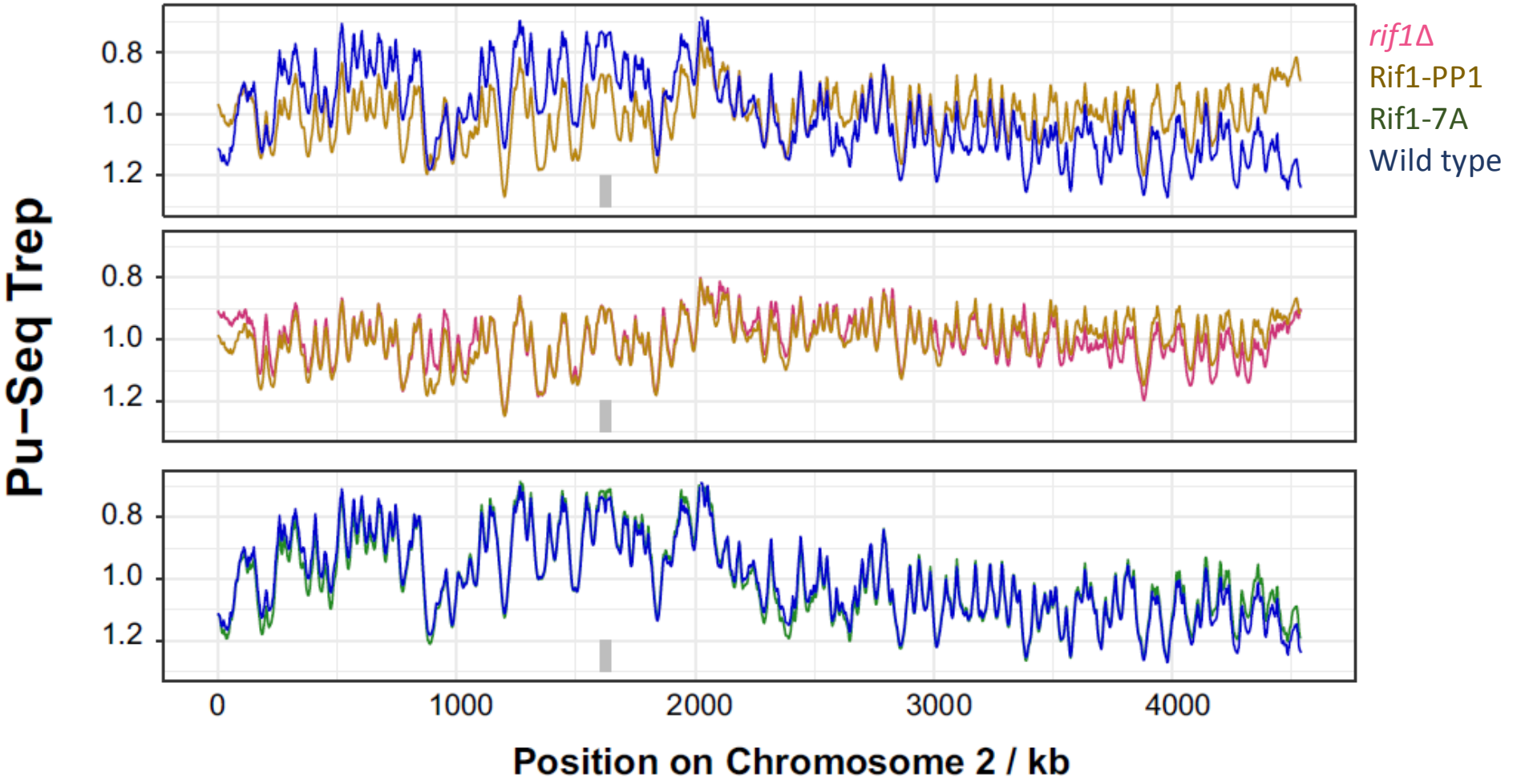
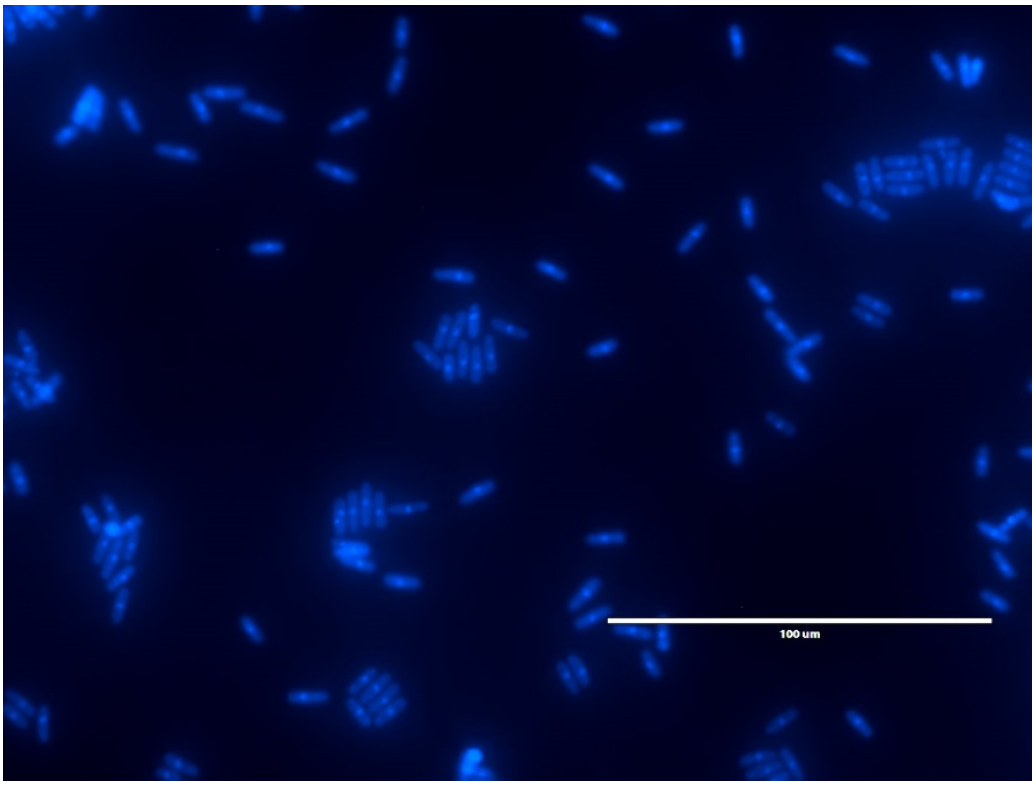

Pu-Seq Trep
0.8
1.0
1.2
0.8
1.0
1.2
0.8
1.0
1.2
0
1000
2000
3000
4000
Position on Chromosome 2 / kb
rif1Δ
Rif1-PP1
Rif1-7A
Wild type

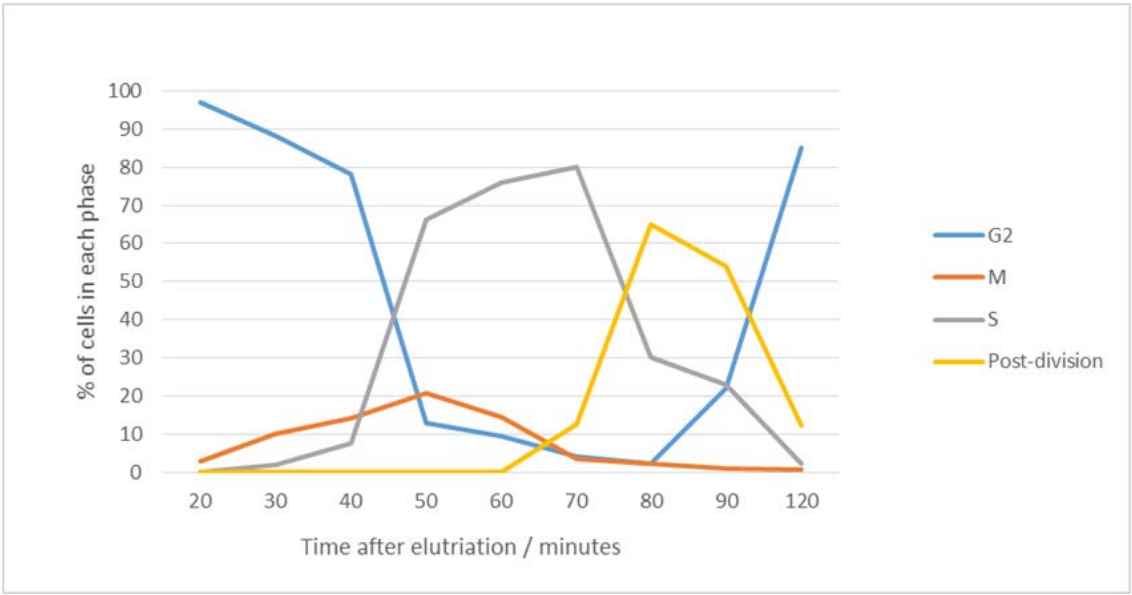
Fig. 5.7 - **Pu-Seq Trep across Chromosome 2 for Rif1-PP1, Rif1-7A, *rif1*Δ and wild type *S. pombe***

Pu-Seq Trep was calculated using the progression of leftward moving forks and assuming a constant fork velocity of 1.5 kb/min. It represents when in S-phase each locus is replicated. The y-axis was, therefore, inverted to maintain the convention of early and late replicating regions being shown on the top and bottom of the graph, respectively. The centromere is marked in gray. The wild type Pu-Seq Trep profile is derived from an average of data generated by 5 wild type Pu-Seq experiments.

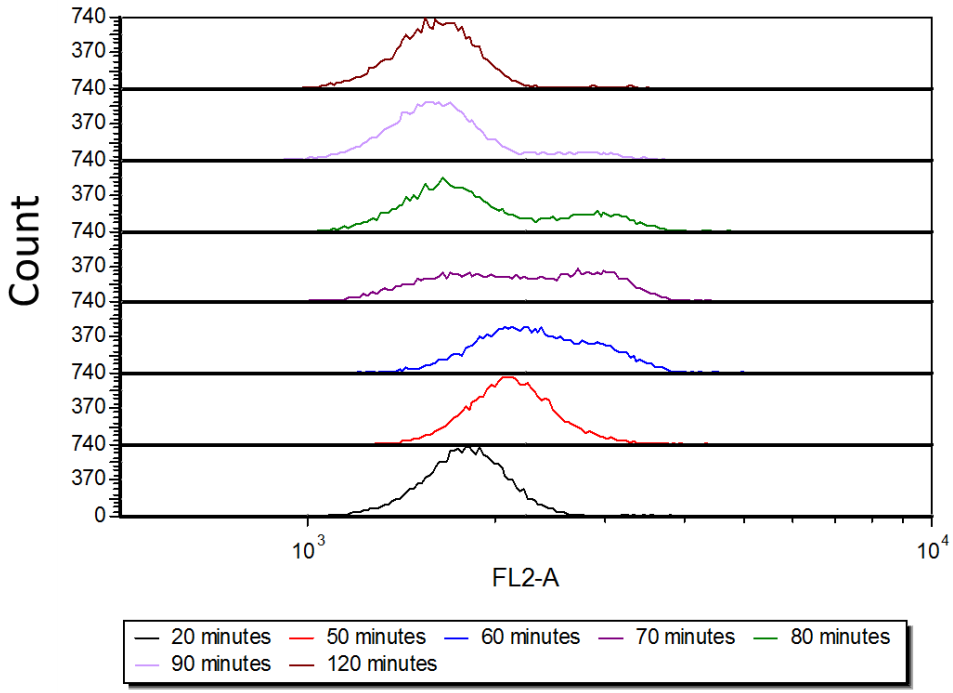
A) Synchronisation after 20 minutes



B) Septation index



C) FACS



D) Elutri-Seq

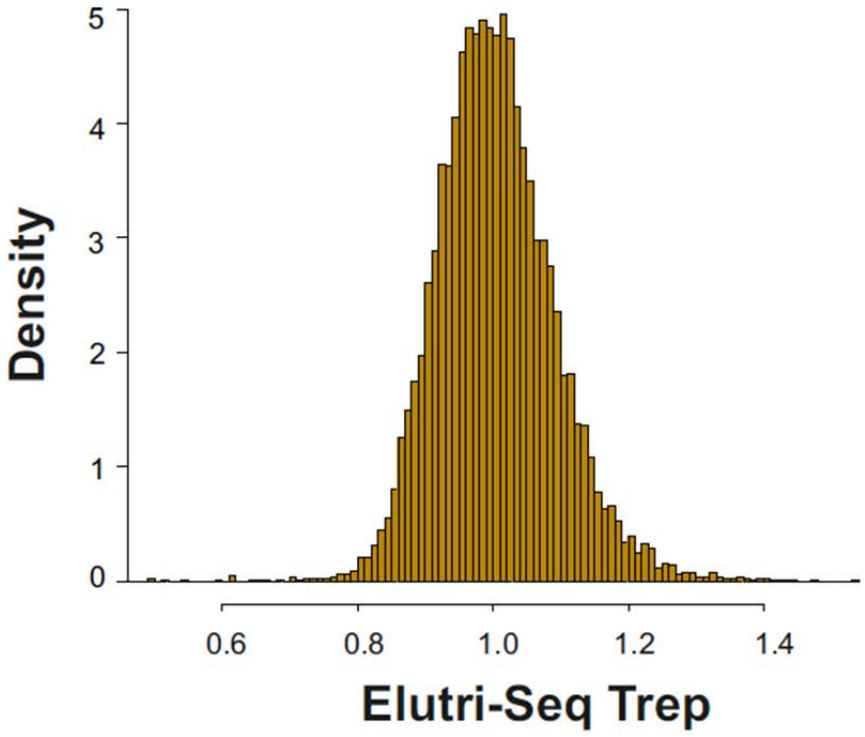


Fig. 5.8 - **Analysis of the cell cycle of Rif1-PP1 cells after elutriation and Rif1-PP1 Elutri-Seq**

A) Synchronisation of Rif1-PP1 20 minutes after elutriation. Cells were stained with 1 µg/mL 4',6'-diamidino-2-phenylindole (DAPI) and 2.5% v/v calcofluor-white (which stain nucleic acid and the septum, respectively). Cells were visualized using an inverted fluorescence microscope (EVOS™ FL).

B) The synchronous passage of Rif1-PP1 cells through the cell cycle after elutriation. Cells were stained and visualised at each time point (as described in A). Cells were counted and divided into different phases of the cell cycle based on their morphology - cells with one nucleus - G_2 ; cells with two nuclei - M-phase ; cells with a septum- S-phase ; two cells joined, without a visible septum, and with one nucleus each - post-division.

C) FACS analysis of cells at each time point. $1.25x10^6$ cells were collected at each time point, stained with propidium iodide and analysed for DNA content on BD Accuri™ C6 Plus flow cytometer.

D) Density distribution of Elutri-Seq Trep for Rif1-PP1 *S. pombe*. Reads from S-phase and G_2 libraries (sequenced in for a coverage over 6 X) were aligned to the reference SP2 genome, binned into 1 kb windows, the number of 5' ends of reads was counted in each bin and normalised to the total number of counts. The ratio the S-phase and G_2 counts in every bin was calculated to produce Elutri-Seq Trep. Rif1-PP1 Elutri-Seq Trep (yellow) is normally distributed, showing that all regions have an equal probability of being replicated at any point during S-phase in Rif1-PP1

5.2.3 Comparing origin activity in late and early replicating regions

To determine whether efficiently and inefficiently firing origins cluster in early and late replicating regions, respectively in *S. pombe* as would have been predicted by the (Yang, Rhind et al. 2010) *S. cerevisiae* model, and whether that is affected in *rif1*Δ and Rif1-PP1, we compared the origin activity in late and early replicating regions.

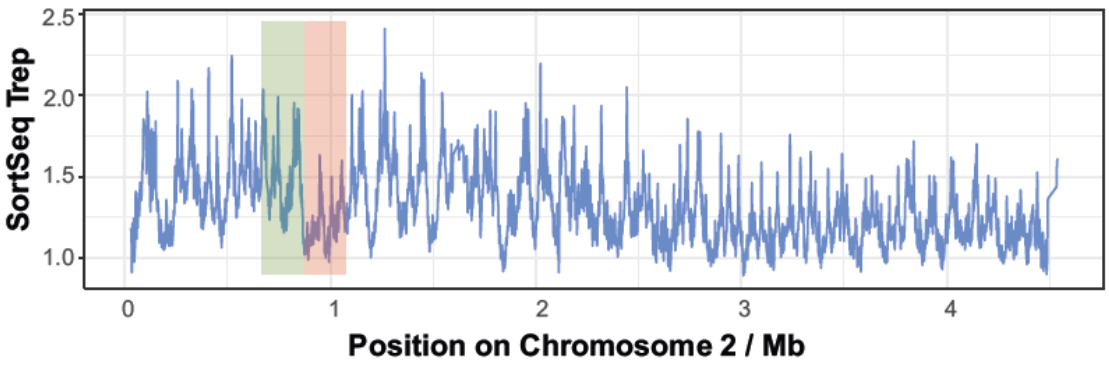
First, we compared origin activity in early and late replicating regions that are directly adjacent to one another (a timing transition zone) - Fig. 5.9 A. In wild type cells, the origin activity in the late region was considerably lower than that in the early region (Fig. 5.9 B), which is in line with the *S. cerevisiae* model (Yang, Rhind et al. 2010). Fig. 5.9 C (panel 1) shows the origin activity in the late and early replicating regions in wild type. The replication forks originating in the early region (marked in green on Fig. 5.9 A) passively replicated the late replicating region (marked in red on Fig. 5.9 A) before the origins in the late region could fire.

Compared to wild type, origins in the late replicating region in *rif1*Δ and Rif1-PP1 fired more efficiently (Fig. 5.9 C), increasing the average origin activity in the late replicating region (Fig. 5.9 B). This change in the probability of origin firing in the late region caused a loss of the RT pattern across the whole region in *rif1*Δ and Rif1-PP1 (Fig. 5.7).

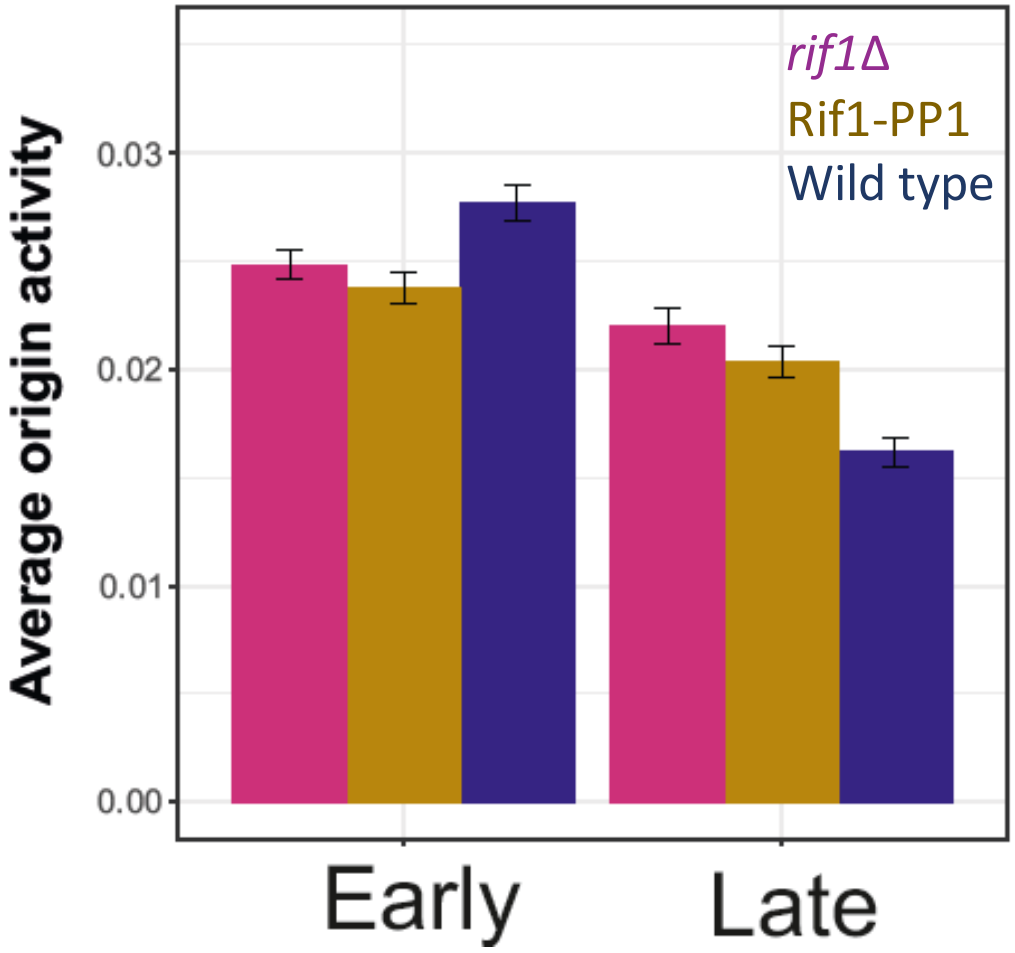
Next, we tested whether early replicating regions contain more efficiently firing origins than late replicating regions globally. We compared the origin firing efficiencies in all the early replicating regions (shown in green on Fig. 4.9) to all the origin firing efficiencies in the late replicating regions. Firstly, in wild type cells, the origin activity in the early replicating regions was only marginally greater than that in late replicating regions (Fig. 5.10), suggesting that both late and early replicating regions are replicated by efficiently and inefficiently firing origins. The origin activity in early replicating regions in *rif1*Δ and Rif1-PP1 was slightly lower, compared to that in wild type (Fig. 5.10). There was, however, no compensatory increase of origin firing in the late regions, as was shown in Fig. 5.9 B.

Taken together, the data suggest that the global RT program in wild type *S. pombe* is not caused by the underlying origin firing efficiencies in the early and late replicating domains.

A) Timing transition zone



B) Average origin activity in late and early regions



C) Origin activity across transition zone

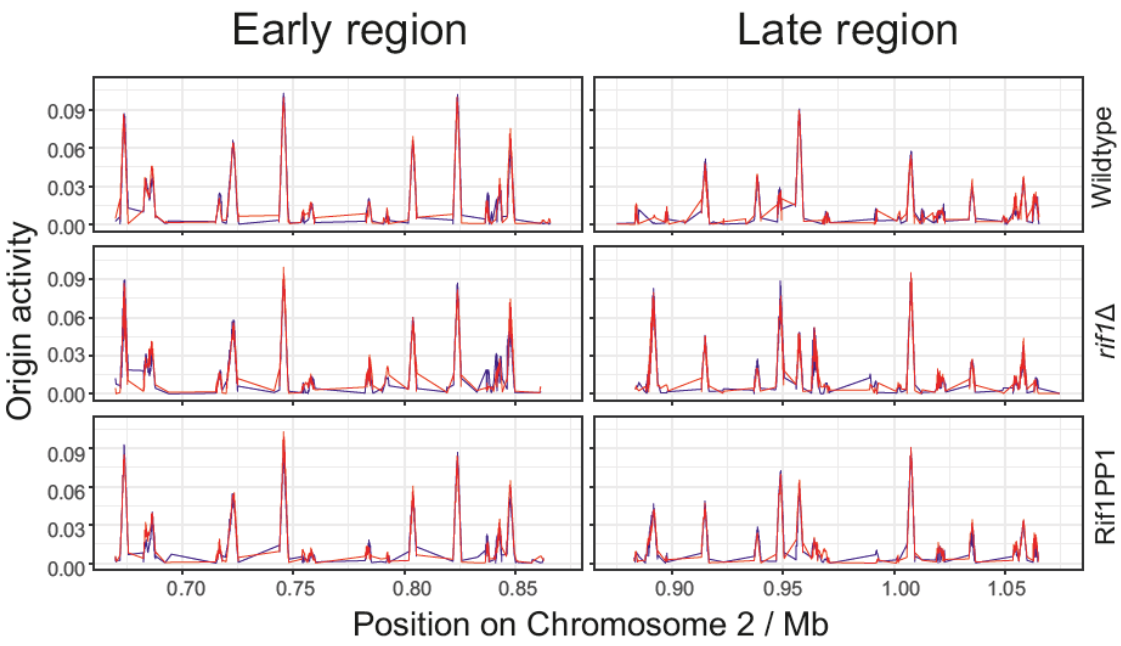


Fig. 5.9 – **Origin activity at an early to late RT transition zone on Chromosome 2 in *rif1*Δ, Rif1-PP1 and wild type backgrounds**

A) Wild type Sort-Seq RT profile across Chromosome 2. Cells were sorted by FACS from an asynchronous culture and sequencing data from synchronous S-phase and G_2 populations were compared to produce a Sort-Seq copy number profile. The data were collected, analysed published in (Daigaku, Keszthelyi et al. 2015). The early to late transition used in the analysis is marked in green and red.

B) The average origin activity in the early (geen) and late (red) replicating region. The average origin activity was calculated by averaging the positive values of the differentials of the polymerase δ usage on the reverse strand and polymerase ε forward strand in all the bins spanning the late and early regions. The error bars represent the standard error of the mean.

C) The origin activity in the early and late replicating regions. Differentials of the polymerase δ usage on the reverse strand (shown in blue) and polymerase ε forward strand (shown in red) in all the bins spanning the late and early regions.

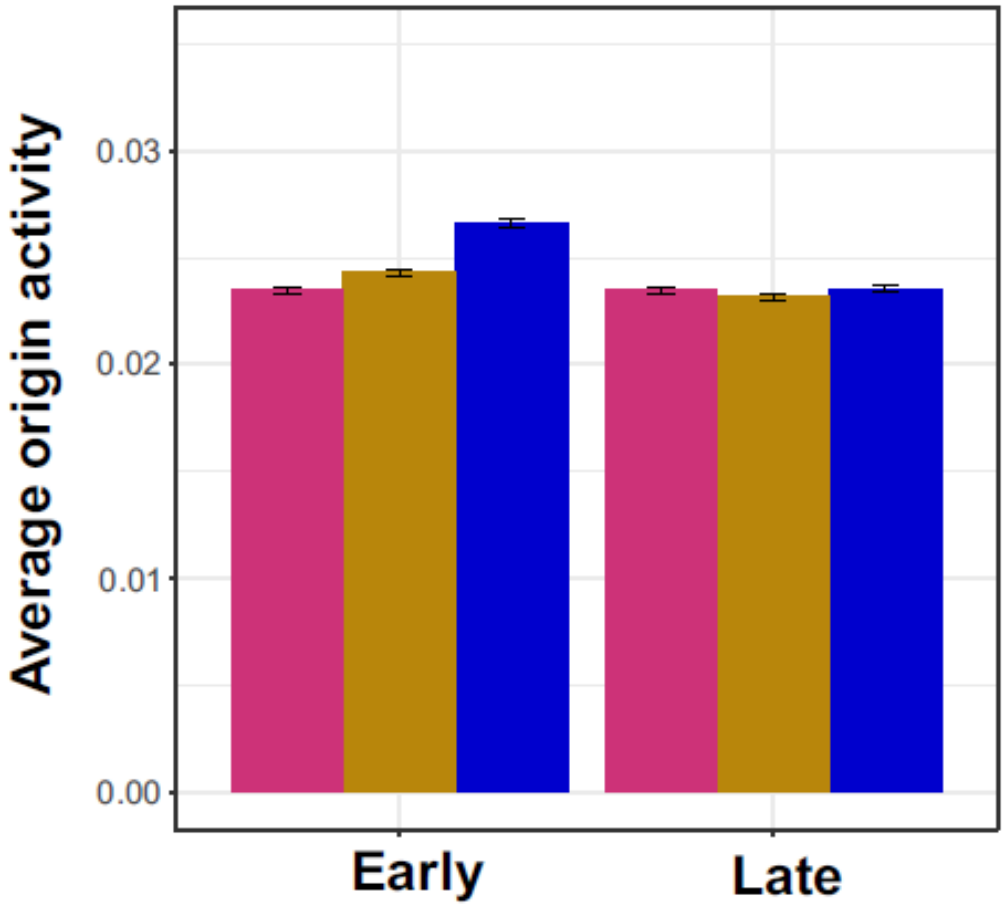


Fig. 5.10 - **The average origin across all late and early replicating regions in *rif1*Δ, Rif1-PP1 and wild type backgrounds**

The average origin activity was calculated by averaging the positive values of the differentials of the polymerase δ usage on the reverse strand and polymerase ε forward strand in all the bins spanning the late and early regions. Early replicating regions in *S. pombe* are marked in green on Fig. 4.9. The error bars represent the standard error of the mean.

5.3 Discussion

In order to validate the RT phenotype described for *rif1*Δ in Chapter 4, we used an independent deep-sequencing approach, called Elutri-Seq. Elutri-Seq combines aspects of Sort-Seq and MFA (discussed in 5.1.3), both of which were used successfully to generate *S. pombe* RT profiles (Daigaku, Keszthelyi et al. 2015). Elutri-Seq was used to create accurate RT profiles (Fig. 5.4) in a more cost and time effective manner.

We generated a wild type Elutri-Seq RT profile from previously published data generated for *S. pombe* MFA (Daigaku, Keszthelyi et al. 2015). The reads from the S-phase and G_2 libraries were aligned to the SP2 reference genome and a ratio between the counts from the two samples was calculated. The wild type Elutri-Seq RT profile had a lower resolution than the Sort-Seq profile (Daigaku, Keszthelyi et al. 2015) but was still an accurate description of the global RT program (Fig. 5.4 B), providing proof of concept.

Elutriation of *rif1*Δ resulted in very tight synchrony (Fig. 5.2). The progression through the cell cycle (Fig. 5.2) was comparable with that reported for wild type (Daigaku, Keszthelyi et al. 2015). Similarly to wild type, *rif1*Δ S-phase and G_2 samples were collected and used to construct Illumina libraries. The *rif1*Δ Elutri-Seq showed a complete loss of the global RT program (Figs. 5.3 and 5.4). No regions of the genome replicated earlier or later than others, i.e., all areas of the genome had an equal probability of being replicated at any point during S-phase in *rif1*Δ. This conclusively verified the *rif1*Δ Pu-Seq Trep phenotype shown in Fig. 4.12.

In *S. cerevisiae*, Rif1 has been proposed to inhibit origin firing in G_1 by recruiting the PP1 phosphatase Glc7 to de-phosphorylate Mcm2-7. It was suggested that this effect was alleviated in S-phase by the phosphorylation of Rif1 by CDK and DDK (Hiraga, Alvino et al. 2014). To determine whether the role of Rif1 in the maintenance of the global RT program in *S. pombe* relied on its ability to interact with phosphatases and/or be phosphorylated, we carried out Pu-Seq analyses in Rif1-PP1 and Rif1-7A strains. These alleles of Rif1 cannot recruit PP1 phosphatases and cannot be phosphorylated, respectively (Fig. 5.5 A). While both alleles had a

wild type distribution of origin firing efficiencies (Fig. 5.5 B), only Rif1-PP1 showed a loss of RT program phenotype similar to that seen in *rif1*Δ (Fig. 5.7). The Rif1-PP1 RT phenotype was verified using Elutri-Seq (Fig. 5.8).

Two conclusions about the global RT program in *S. pombe* can be drawn from these data. Firstly, the maintenance of the global pattern is reliant on the action of PP1 phosphatases recruited by Rif1. Secondly, the global RT program is not caused by the underlying origin firing efficiencies in early and late replicating domains, as has been hypothesised in *S. cerevisiae* (Yang, Rhind et al. 2010). Therefore, despite the fact that the local effects of Rif1 on origin activity are also driven by its interaction with phosphatases (Fig. 5.6), we propose that the global RT program and the global origin firing efficiencies are unlinked in *S. pombe*.

Instead, we propose a model where Rif1 mediated de-phosphorylation plays a role in establishing early and late S-phase. In our model, during early S-phase, origins in late replicating regions would be inhibited from firing (possibly through the direct action of PP1 phosphatases). In late S-phase, this inhibition would be lifted and origins in late regions would fire equally efficiently to their counterparts in early replicating regions. This model would explain the change in the relative RT of subtelomeres in *rif1*Δ (Fig. 4.12) and *taz1*Δ (Fig. 4.10 A) without an increase in the origin activity in the region (Fig. 4.7).

Our data, which compared the origin activity globally in early and late replicating regions support this model (Fig. 5.10). These data show that in wild type cells origins in both early and late replicating regions fire with very comparable efficiencies. Furthermore, in *rif1*Δ and Rif1-PP1, where the global pattern is lost, origins in late replicating regions did not become more efficiently firing, as would be expected from a model which dictates that the earlier replication of these regions had to be driven by an increase in the underlying origin activity.

Rif1-PP1 and *rif1*Δ origin activity in early replicating regions was slightly lower than that in wild type (Fig. 5.10). We propose that the lower origin activity is caused by a decrease in the availability of limiting factors. Given that in *rif1*Δ and Rif1-PP1, all origins have an equal probability of firing, the relative efficiency with which the previously "efficient" origins fired, decreases.

We cannot, however, rule out the possibility that the marginal difference between origin activity in late and early replicating regions seen in wild type cells (Fig. 5.10) could be responsible for generating the global RT program. Similarly, it is also possible that the small decrease in origin activity in early replicating regions in *rif1*Δ and Rif1-PP1 (Fig. 5.10) can account for the loss of the global RT. To irrefutably assert that the global RT program in *S. pombe* is not driven by relative differences in origin activity, we propose the building of a new forward mathematical model of DNA replication, where origin activity and replication timing are unlinked. The model could be used to predict *in silico* the extent to which the origin activity in the early and late replicating domains would have to be affected to produce a "flat" RT profile, as is seen in *rif1*Δ and Rif1-PP1 (Fig. 5.7).

We speculate that the "switch" between early and late S-phase could be caused by the phosphorylation of Rif1. Although our data do not show a change in RT for Rif1-7A (Fig. 5.7), it has been previously reported that the allele does not show stronger interactions with the phosphatases (Dave, Cooley et al. 2014) and, therefore, may not completely inhibit Rif1 phosphorylation. We propose that there may be a threshold of Rif1 phosphorylation in mid to late S-phase at which the interactions between Rif1 and the PP1 phosphatases are inhibited. This would stop the de-phosphorylation of Mcm2-7 and allow the efficient firing of origins in late replicating regions, during late S-phase.

A very similar analysis of the effect of Rif1 on the origin activity and global RT program in *S. cerevisiae* has been recently published (Hafner, Lezaja et al. 2018). While (Hafner, Lezaja et al. 2018) show a similar Rif1 (and PP1) dependent inhibition of origin firing, the global RT profiles (generated using Sort-Seq) were unaffected in *rif1*Δ. Instead, the pattern of Sort-Seq copy number "peaks" around origins of replication were flatter. This further reinforces the idea that while the global RT program in *S. cerevisiae* may be driven by changes in origin activity, this may not be descriptive of replication dynamics in other eukaryotes.

Chapter 6

6.1 Introduction

To further explore the underlying mechanisms, we investigated the role played by the nuclear distribution of chromatin and effect of Mrc1 on the establishment of the global RT program

Firstly, we analysed to nuclear localisation of Rif1 to determine whether it may play a role in the tethering of chromatin to the nuclear periphery. Secondly, to establish whether a change in the nuclear localisation of a genomic region could impact its relative RT, we generated Pu-Seq RT profiles from cells with an abrogated attachment of telomeres to the nuclear periphery.

Finally, we examined the involvement of Mrc1, whose deletion has been previously described to have similar effects on global origin firing to that described for *rif1*Δ (Hayano, Kanoh et al. 2011, Hayano, Kanoh et al. 2012)

6.1.1 Linking the nuclear distribution of chromatin and the replication timing program

As discussed in 1.4.2, the boundaries of metazoan TADs, LADs and replication timing domains are significantly correlated, suggesting a strong link between the nuclear distribution of chromatin and the global RT program (Rivera-Mulia and Gilbert 2016). Comparison of human and murine RT profiles across regions of synteny, showed a considerable conservation of replication timing domains (Ryba, Hiratani et al. 2010), suggesting an important evolutionary role. No link between nuclear architecture and replication timing domains has been yet reported in *S. pombe*. It is important to note that the genome organisation is different between yeast and metazoans, e.g., neither *S. cerevisiae* nor *S. pombe* genomes organise into metazoan-like TADs (Duan, Andronescu et al. 2010, Mizuguchi, Fudenberg et al. 2014, Hsieh, Weiner et al. 2015, Eser, Chandler-Brown et al. 2017). *S. pombe* "globules" do, however, share some similarities with TADs. For example, the formation of globules is dependent on cohesin and independent of heterochromatin (Mizuguchi, Fudenberg et al. 2014, Pichugina, Sugawara et al. 2016). The molecular conservation suggests that the globules and TADs may be related and the nuclear distribution of chromatin in *S. pombe* may play a role in the global RT program.

It was recently suggested that in mESCs, the nuclear architecture is organised by Rif1 (Foti, Gnan et al. 2016). Despite the widespread role of Rif1 in global RT

program (Cornacchia, Dileep et al. 2012, Hayano, Kanoh et al. 2012, Peace, Ter-Zakarian et al. 2014, Kanoh, Matsumoto et al. 2015), no connection between Rif1 and the organisation of chromatin domains has been suggested in yeast.

In *S. cerevisiae* Rif1 requires Pfa4 dependent palmitoylation to associate with the nuclear periphery (Park, Patterson et al. 2011). The loss of Rif1 palmitoylation, however, does not lead to a change in the global RT program of *S. cerevisiae* (Peace, Ter-Zakarian et al. 2014), suggesting that although Rif1 is located at the nuclear periphery and may play some role in tethering chromatin it, the Rif1 dependent chromatin distribution is not necessary for the global RT program in *S. cerevisiae.*

A quantitative analysis of the *S. pombe* proteome, estimated only ~135 Rif1 molecules per cell, during vegetative growth (Marguerat, Schmidt et al. 2012). It has also been reported that endogenous levels of Myc-GFP tagged Rif1 yield insufficient fluorophore signal to detect using wide field fluorescence microscopy (Zaaijer, Shaikh et al. 2016). Overexpression of N-terminally tagged Rif1-GFP resulted in a "nuclear haze", with foci co-localising with Taz1. Overexpressed Rif1-GFP was also reported to localise to anaphase bridges. The action of Rif1 on anaphase bridges, however, was shown to be independent of its role in S-phase (Zaaijer, Shaikh et al. 2016) and will not be discussed here. Additionally, instead of forming extensive Rif1 associated domains (RADs), as was reported in metazoans (Foti, Gnan et al. 2016), the binding of Rif1 in *S. pombe* was detected at only discreet 90 sites along the genome (Kanoh, Matsumoto et al. 2015). These data suggest that Rif1 may not play a significant role in establishing the nuclear distribution of chromatin in *S. pombe*.

Nevertheless, a potential Rif1-independent tethering of chromatin to the nuclear periphery could be important for the maintenance of the global RT program in yeast. When considering the potential nuclear organisation underlying the global RT in *S. pombe*, differences between yeast and metazoan nuclei must be considered. For example, nuclear laminas are not formed in yeast and chromatin, therefore, does not organise into true metazoan-like LADs. A comparative sequence analysis of 28 eukaryotic genomes, however, identified two proteins in *S. pombe* with considerable sequence similarity to the LEM (Lap2, Emerin, Man1)

family of LAP (lamina associated polypeptides) proteins (Mans, Anantharaman et al. 2004). These two proteins, Lem2 and Man1, have been shown to anchor telomeres to the nuclear periphery (Gonzalez, Saito et al. 2012). Unlike Lem2, however, Man1 does not play a role in the maintenance of nuclear stability (Gonzalez, Saito et al. 2012). This suggests that Man1 may act as a platform to allow the binding of chromatin and, potentially other proteins, to the nuclear envelope, similarly to the metazoan lamina. Consistent with this, almost a third of the *S. pombe* genome was shown to interact with Man1 (Steglich, Filion et al. 2012), which is comparable to the 40% of the mouse and human genomes found in LADs (Guelen, Pagie et al. 2008, Peric-Hupkes, Meuleman et al. 2010). Similarly to loci found in metazoan LADs, Man1 associated loci were significantly enriched in weakly expressed genes and low RNA polymerase II levels (Steglich, Filion et al. 2012). Taken together, these data suggest that Man1 could have a lamin-like function in *S. pombe* and contribute to the distribution of chromatin around the nucleus. This chapter will explore the impact of nuclear distribution of chromatin on RT in *S. pombe*, by studying the effect of *man1*Δ on the global RT program. Studying the effects of disrupting the cohesion dependent globules (Mizuguchi, Fudenberg et al. 2014) is more complex as cohesion null (*rad21*Δ) *S. pombe* mutants are inviable.

6.1.2 Role of Mrc1 in the RT program of *S. pombe*

Mrc1 (mediator of replication checkpoint) was originally described in yeast as a signal transducer of the DNA replication checkpoint (Alcasabas, Osborn et al. 2001, Tanaka and Russell 2001). As discussed in 1.2.3.2, Mrc1 travels with the replisome and interacts with the FPC to maintain genome integrity in the presence of replication stress (Katou, Kanoh et al. 2003, Tourriere, Versini et al. 2005). Additionally to its role in genome stability, Mrc1 was suggested to play a role in the regulation of origin firing (Hayano, Kanoh et al. 2011).

In *S. pombe*, Mrc1 binds to a number of origins, many of which have been reported to be early/efficiently firing, in a Cdc45 and Hsk1 independent manner (Hayano, Kanoh et al. 2011). Given that Mrc1 interacts directly with Hsk1 (Shimmoto, Matsumoto et al. 2009), it was speculated that Mrc1 could recruit Hsk1 to efficiently firing origins (Hayano, Kanoh et al. 2011). Interestingly, however, the

firing of the origins bound by Mrc1 is stimulated in *mrc1*Δ (Hayano, Kanoh et al. 2011), suggesting that Mrc1 has an inhibitory effect on origin firing.

It has been recently shown that an intramolecular interaction of Mrc1 (regulated by Hsk1 phosphorylation) drives the inhibitory effect of Mrc1 on origin firing (Matsumoto, Kanoh et al. 2017) (Fig. 6.1). Mrc1 interacts directly with Hsk1 via its Hsk1 bypass segment (HBS). The intramolecular interaction between the HBS and the N-terminal target of HBS (NTHBS) induces an Mrc1 conformation in which it can bind to early firing origins and inhibit their firing ("brake-on" conformation). Upon interacting with Hsk1, Mrc1 can be phosphorylated on a region adjacent to the NTHBS, which interrupts the intramolecular interaction and induces a "brake-off" conformation, which does not inhibit origin firing (Matsumoto, Kanoh et al. 2017).

An allele of Mrc1 that cannot form the intramolecular interaction with NTHBS, resulting in constitutive "break off" configuration (Mrc1ΔHBS) was shown to advance the firing of late/inefficient origin *ars1*, similarly to *mrc1*Δ. Unlike *mrc1*Δ, however, Mrc1ΔHBS was not sensitive to genotoxic agents (Matsumoto, Kanoh et al. 2017). Conversely, a checkpoint deficient mutant of Mrc1 (Mrc1-3A) was sensitive to HU but did not stimulate origin firing (Hayano, Kanoh et al. 2012, Matsumoto, Kanoh et al. 2017). These data show that the inhibition of origin firing at Mrc1 bound origins is independent of its role in the intra-S-phase checkpoint (Hayano, Kanoh et al. 2011, Matsumoto, Kanoh et al. 2017).

A model has been proposed where Mrc1 acts in concert with Rif1 to inhibit origin firing in early and late replicating regions, respectively, setting up the global RT program (Masai, Yang et al. 2017). In this model, Mrc1 binds to early/efficiently firing origins in the "break on" configuration, preventing their precocious firing in a checkpoint independent manner. Upon binding of Hsk1, Mrc1 is phosphorylated and changes configuration to "break off". As a result of Mrc1-Hsk1 binding, the early/efficiently firing origins are brought into close proximity of Hsk1, which allows their early/efficient firing. The model also proposes that, independently of Mrc1 action at early firing origins, Rif1 mediated de-phosphorylation of Mcm2-7 inhibits the firing of late/inefficient origins. Rif1 and Mrc1 were proposed to be "dual regulators" of origin firing in *S. pombe* (Masai, Yang et al. 2017). Given that

the authors do not distinguish between the efficiency and timing of origin firing, it was proposed that both Mrc1 and Rif1 regulated "origin firing timing" (Masai, Yang et al. 2017), suggesting a possibility that Mrc1 may be involved in the global RT program in *S. pombe*.

Similarly to *rif1*Δ, *mrc1*Δ bypasses the lethality of *hsk1*Δ (Hayano, Kanoh et al. 2011, Hayano, Kanoh et al. 2012). Both Mrc1ΔHBS and Mrc1-3A partially restore viability to *hsk1*Δ (Matsumoto, Kanoh et al. 2017), suggesting that, in the case of *mrc1*Δ, the lethality bypass is reliant on both the checkpoint and non-checkpoint roles of Mrc1. The effects of *mrc1*Δ and *rif1*Δ on *hsk1*Δ lethality bypass are additive, (Hayano, Kanoh et al. 2012), showing that they work in independent pathways.

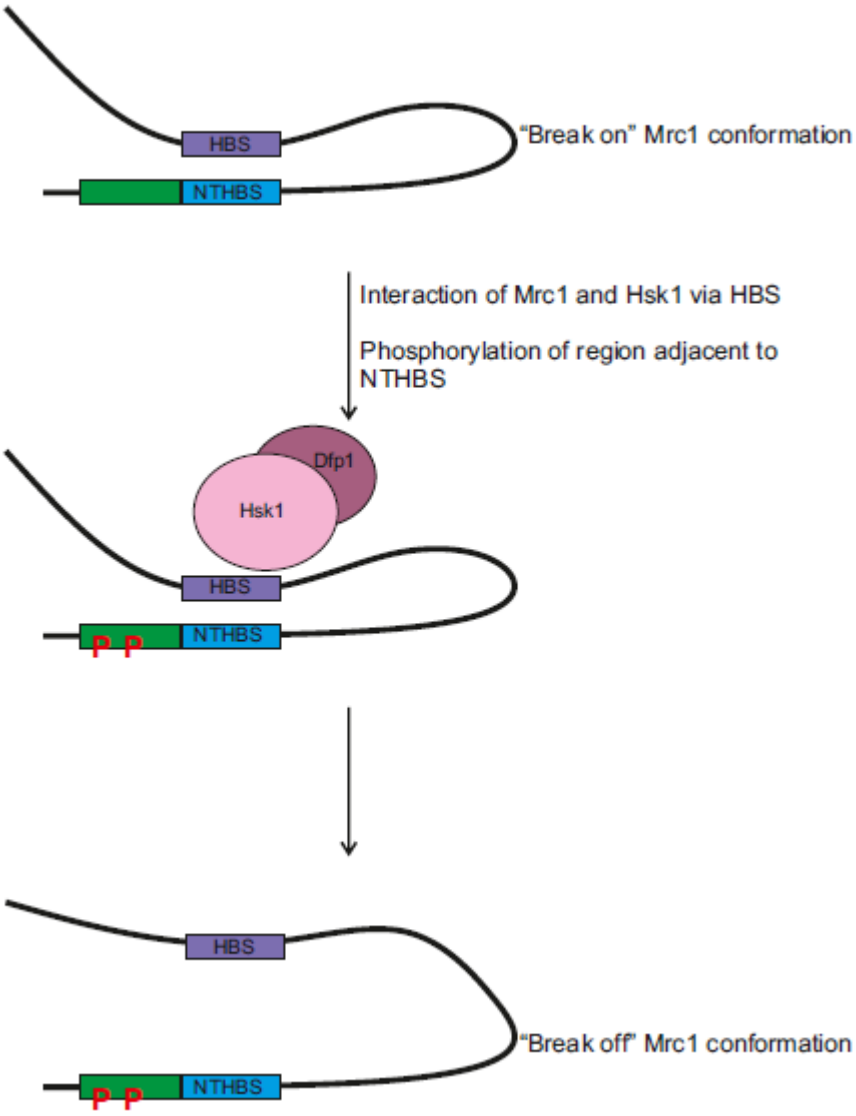


Fig. 6.1 - **Regulation of Mrc1 by Hsk1 phosphorylation**

The HBS (Hsk1 bypass segment) interacts with the N-terminal target of HBS (NTHBS) inducing a "break on" Mrc1 conformation. In this conformation, Mrc1 binds to early origins and inhibits their firing. Upon interaction with Hsk1 (via its HBS segment) Mrc1 is phosphorylated on a region adjacent to the NTHBS. This disrupts the interaction between the HBS and NTHBS, inducing a "break off" Mrc1 conformation.

Adapted from (Masai, Yang et al. 2017).

6.1.3 Aims

To determine the localisation of endogenously expressed Rif1, we tagged Rif1 with photoactivatable mEOS 3.2 tag and analysed the nuclear distribution using photo-activated localization microscopy. To determine whether the interaction of Rif1 with PP1 phosphatases affects this distribution, we also analysed the localisation of Rif1-PP1.

To assess the impact of disrupting the nuclear distribution of chromatin in the nucleus on the global RT program and/or the relative RT of the subtelomeric regions, we carried out a Pu-Seq analysis of *man1*Δ. We compared the effects of *man1*Δ on global RT to those seen in *rif1*Δ.

Next, to determine whether heterochromatin affects the global RT program, we analysed the RT profiles of heterochromatin null mutants (*clr4*Δ and *swi6*Δ).

Finally, we generated an Elutri-Seq RT profile for *mrc1*Δ to determine whether Mrc1 acts synergistically with Rif1 to establish the global RT program.

6.2 Results

To determine whether Rif1, expressed at endogenous levels, localises to the periphery and whether this localisation is dependent on the interaction with PP1 phosphatases we tagged Rif1 and Rif1-PP1 with photoactivatable mEos3.2. The cells were imaged using photo-activated localization microscopy (PALM), a super-resolution technique that allows the visualisation of single molecules and has been previously used to image DNA bound proteins in *S. pombe* (Etheridge, Boulineau et al. 2014). Unfortunately, there is no single direct orthologue of Pfa4 in *S. pombe* (pers. com. Ulrich Rass). We were, therefore, unable to directly assess whether the localisation of Rif1 in *S. pombe* is dependent on palmitoylation

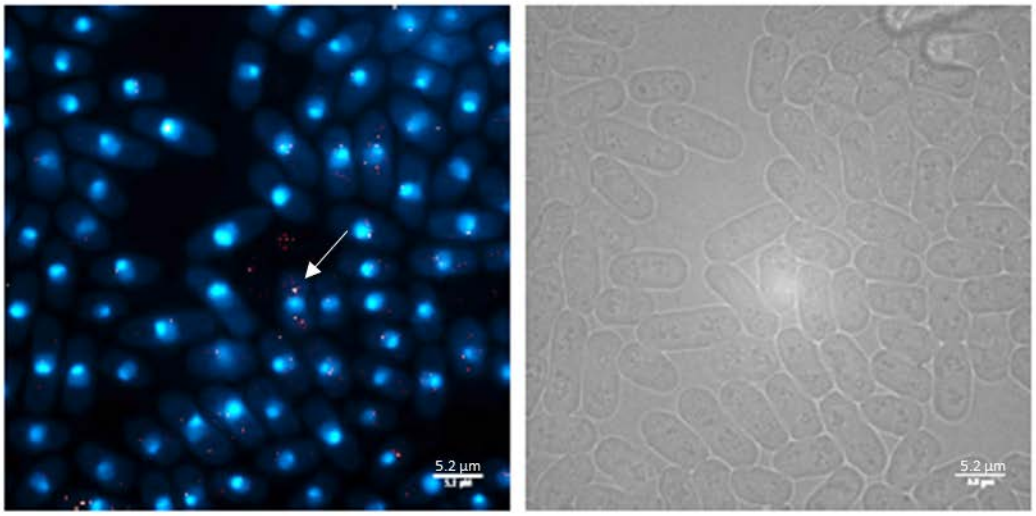
6.2.1 Distribution of Rif1 and Rif1-PP1 in the nucleus

The Rif1 and Rif1-PP1 tagged strains were constructed using recombination mediated cassette exchange (RMCE - described in 2.2.3.3) (Watson, Garcia et al. 2008) and crossed with Gar2-GFP (kindly shared by Dr. Jo Murray). Gar2 is a non-ribosomal protein (Sicard, Faubladier et al. 1998) whose fluorescence marks the nucleolus. The strains were grown to OD 0.5 in EMM and prepared for imaging as

described in (Etheridge, Boulineau et al. 2014). The visualisation and the image processing was carried out by Dr. Thomas Etheridge, as described in 2.2.6.

Surprisingly, only ~20 localisations were seen for both Rif1 and Rif1-PP1 (Fig. 6.2), suggesting that the endogenous levels of Rif1 may not be sufficient to visualise even using super-resolution microscopy. It is possible, however, that the tagging of Rif1 with mEos3.2 affected the expression levels of Rif1 and the levels observed were not truly endogenous. Regardless, due to the low levels of foci observed, Rif1-specific localisation could not be reliably distinguished from false localisations caused by cellular auto-fluorescence. A quantitative comparison of the distributions was, therefore, not possible. It was noted, however, that many localisation in Rif1 and Rif1-PP1 cells, were seen at the nuclear and nucleolar peripheries (Fig. 6.2), suggesting that Rif1 may localise to the periphery, similarly to what has been reported for *S. cerevisiae* (Park, Patterson et al. 2011) and that this localisation does not appear to be driven by the interaction with PP1 phosphatases.

A) mEos3.2-Rif1 Gar2-GFP



B) mEos3.2-Rif1PP1 Gar2-GFP

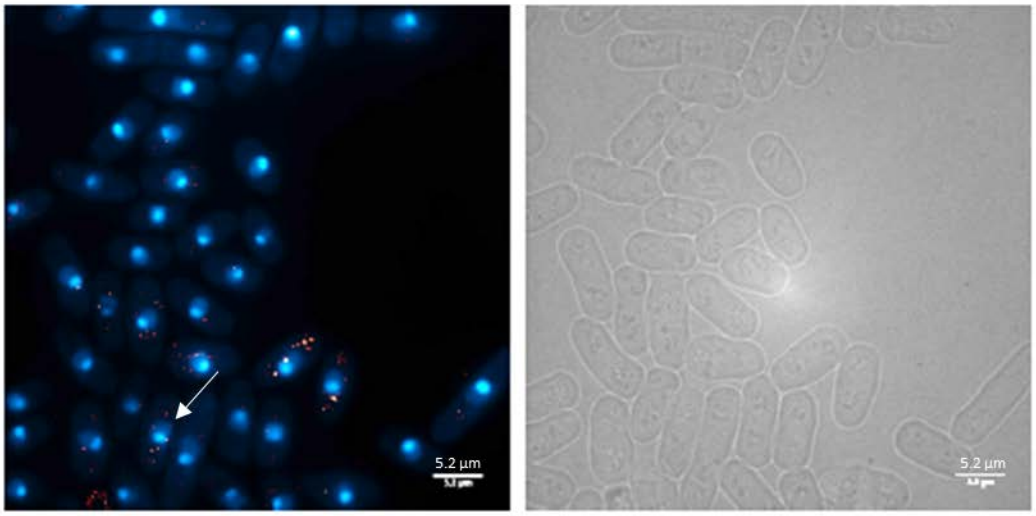


Fig. 6.2 - **Localisation of endogenously expressing mEos3.2 tagged Rif1 and Rif1-PP1 in live *S. pombe***

Gar2-GFP was excited using the 488 nm laser at 15% power and images were acquired using 100 ms exposure time. Rif1-mEos3.2 and Rif1PP1-mEos3.2 were excited using dual continuous 405 nm and 561 nm excitation (1 W/cm^2 and 1 kW/cm^2, respectively) for 1000 frames with 50 ms exposure time. Multiple fields of view were acquired per experimental repeat. A representative frame with a white light image of the same field of view is shown on the right.

A) Overlay image of Gar2-GFP (blue) and mEos3.2-Rif1 localisations (orange).

B) Overlay image of Gar2-GFP (blue) and mEos3.2-Rif1-PP1 localisations (orange).

White arrows point to representative Rif1-mEos3.2 and Rif1PP1-mEos3.2 localisations.

6.2.2 Impact of the nuclear distribution of chromatin on global RT pattern

Next, we went on to determine whether a Rif1-independent disruption of the chromatin tethering to the nuclear periphery affects the global RT program and/or the relative RT of the subtelomeric regions by carrying out a Pu-Seq analysis of *man1*Δ cells.

The polymerase usage data for *man1*Δ was different to the wild type data (Fig. 6.3 A). The range of polymerase usage (on a scale of 0 to 1) was narrower in *man1*Δ compared to wild type (Fig. 6.3 A). The mean usage of polymerase δ on the forward strand ranged from 0.31 to 0.72 for wild type and 0.41 to 0.60 for *man1*Δ. This narrow range was reminiscent of that seen in data from libraries made using Klenow polymerase (polymerase δ usage ranged between 0.36 and 0.65- discussed in 3.2.3, shown on Fig. 3.8). To determine whether the narrow range of polymerase usage seen could have been caused by a lower coverage of the genome in *man1*Δ, we calculated the coverage (using the Lander/Waterman equation (Lander and Waterman 1988)) and compared it to that of wild type (Table 6.1).

Table 6.1 – **Number of reads obtained and mapped for the wild type and *man1*Δ Pu-Seq libraries**

The data shown here were derived from libraries that were pooled and sequenced together on one Illumina flow cell lane. The overall alignment rate represents the number of R1 and R2 mates that aligned in pairs (concordantly and discordantly) and single reads that aligned in single end mode.

Experiment	Replicative polymerase mutated in the strain	Number of reads mapped to the reference genome	Coverage
Wild type	Polymerase δ	13,890,735	55
	Polymerase ε	17,713,962	70
***man1*Δ**	Polymerase δ	15,032,380	60
	Polymerase ε	13,208,267	52

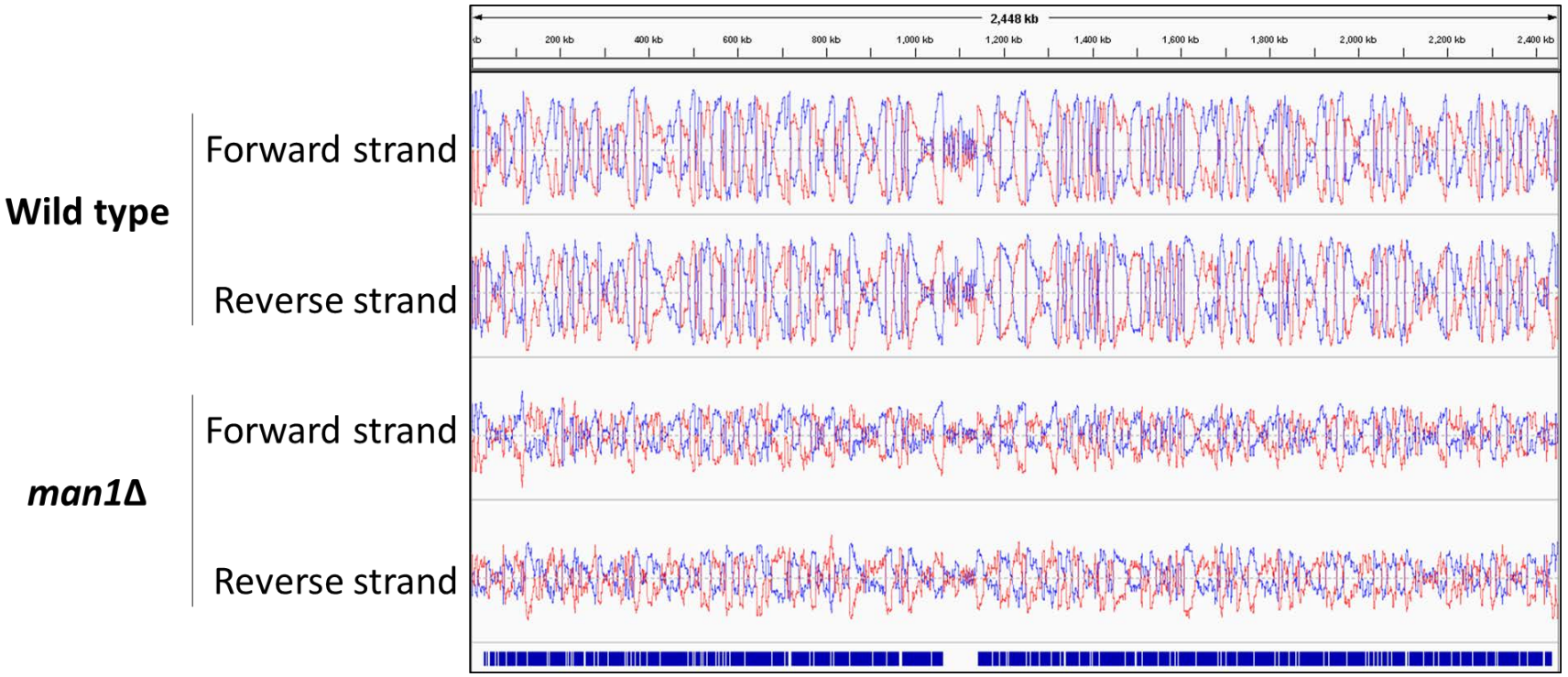
The deviation between the coverage of the polymerase δ and ε libraries was marginally greater than that seen in the wild type T4 libraries described in Chapter 3 (Table 3.1). The overall coverage, however, was comparable both within and

between the experiments. It is, therefore, possible that the differences in polymerase usage seen in *man1*Δ (Fig. 6.3 A) are the result of a biological phenomenon. Given, however, that we have not yet repeated the *man1*Δ Pu-Seq experiment, it is impossible to rule out a technical issue.

The *man1*Δ global RT profile was different from that of both wild type and *rif1*Δ (Fig. 6.3 B). In *man1*Δ, the wild type transition zone between early and late replication in the middle of Chromosome 2 and the dip in the regional RT on the right arm of Chromosome 1 were present. The transitions were, however, less pronounced than in wild type. It is possible that the sharpness of the transitions were masked by the potential technical error that caused the low polymerase usage ratios. We, therefore, compared the *man1*Δ Pu-Seq Trep profile to a Pu-Seq Trep profile derived from the Klenow data described in Table 3.1. The narrow range of polymerase usage values affects the Pu-Seq Trep profile but cannot fully account for the changes seen in *man1*Δ (Fig. 6.3 C). A comparison with the Klenow wild type data suggests that only the relative replication timing of the subtelomeres and ~1 Mb long region on the right arm Chromosome 1 were affected in *man1*Δ (Fig. 6.3 C). The region on Chromosome 1 replicated later in *man1*Δ (6.3 C), compared to wild type. In contrast, subtelomeres replicated early in *man1*Δ (6.3 C), a phenotype reminiscent of *taz1*Δ (Fig. 4.10 A) and *rif1*Δ (Fig. 4.12).

The distribution of the normalised origin firing efficiencies in *man1*Δ was not bimodal (Fig. 6.4 A). Compared to wild type (including a distribution of origin firing efficiencies derived from wild type Klenow data), substantially more origins fired inefficiently (i.e., with 10% to 40% firing efficiency). The few efficiently firing origins did not cluster in late or early replicating regions- there was no difference between the average origin activity in *man1*Δ in regions that replicated early and late in wild type (Fig. 6.4 B). The absolute levels of the average origin activity shown in Fig. 6.4 B were lower for *man1*Δ than wild type or *rif1*Δ. The low ratios of polymerase usage, however, would directly affect these values. Given that they could be the result of a technical error, we cannot analyse the difference in the absolute values of the origin activities between wild type and *man1*Δ.

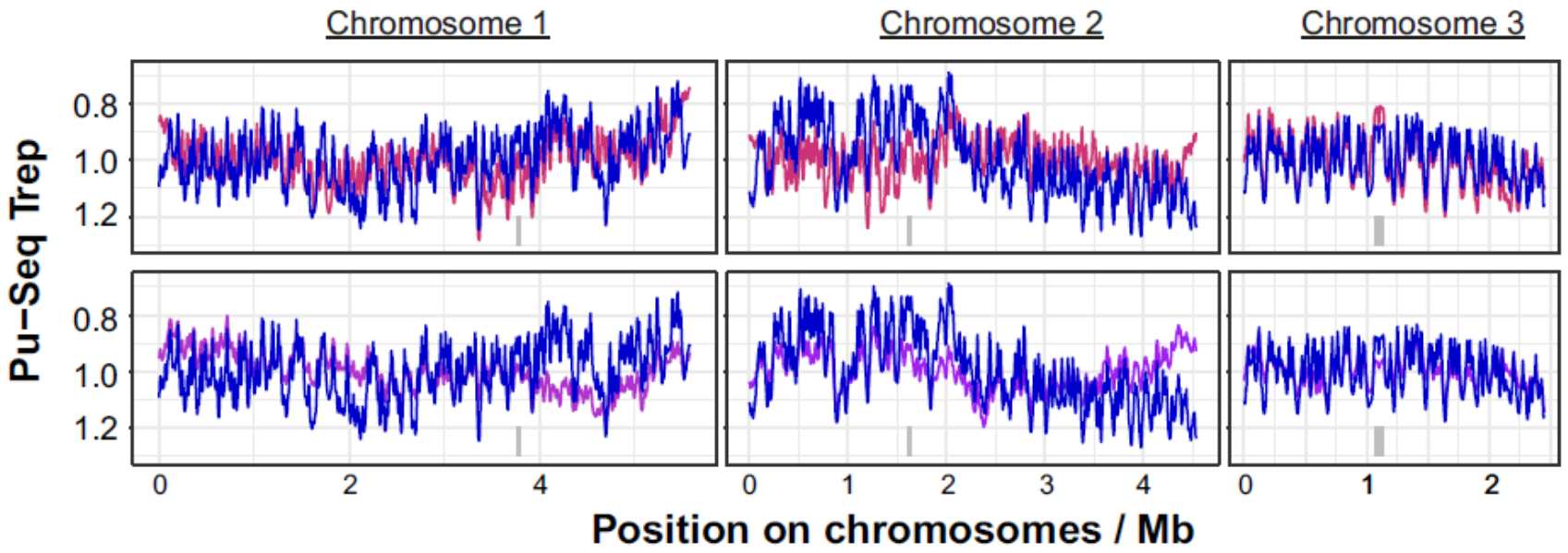
A) Polymerase usage in *man1Δ*



B-C) Pu-Seq Trep in *man1Δ*

rif1Δ
man1Δ
Wild type

B)



C)

rif1Δ
Wild type (Klenow)

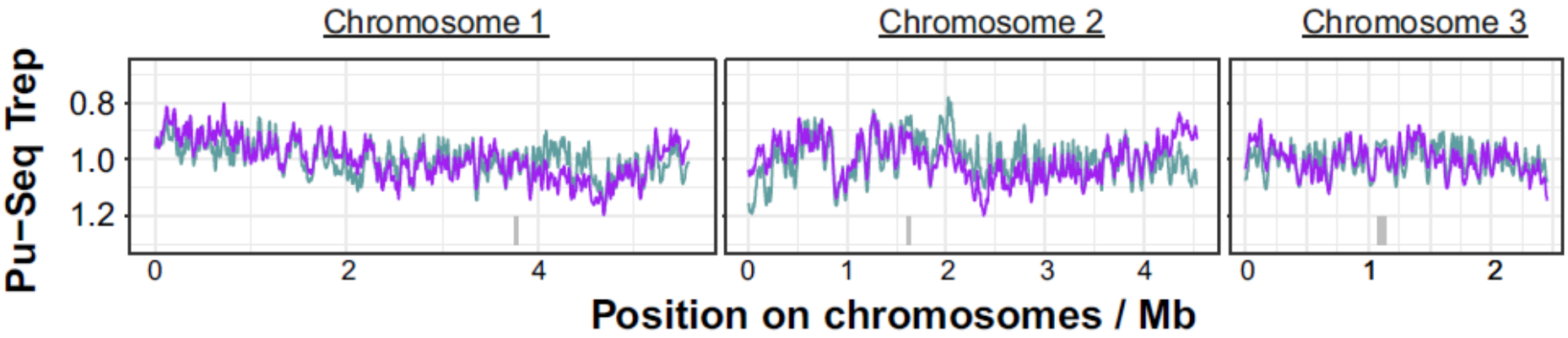


Fig. 6.3 – **Polymerase usage and Pu-Seq Trep in *man1*Δ and wild type *S. pombe***

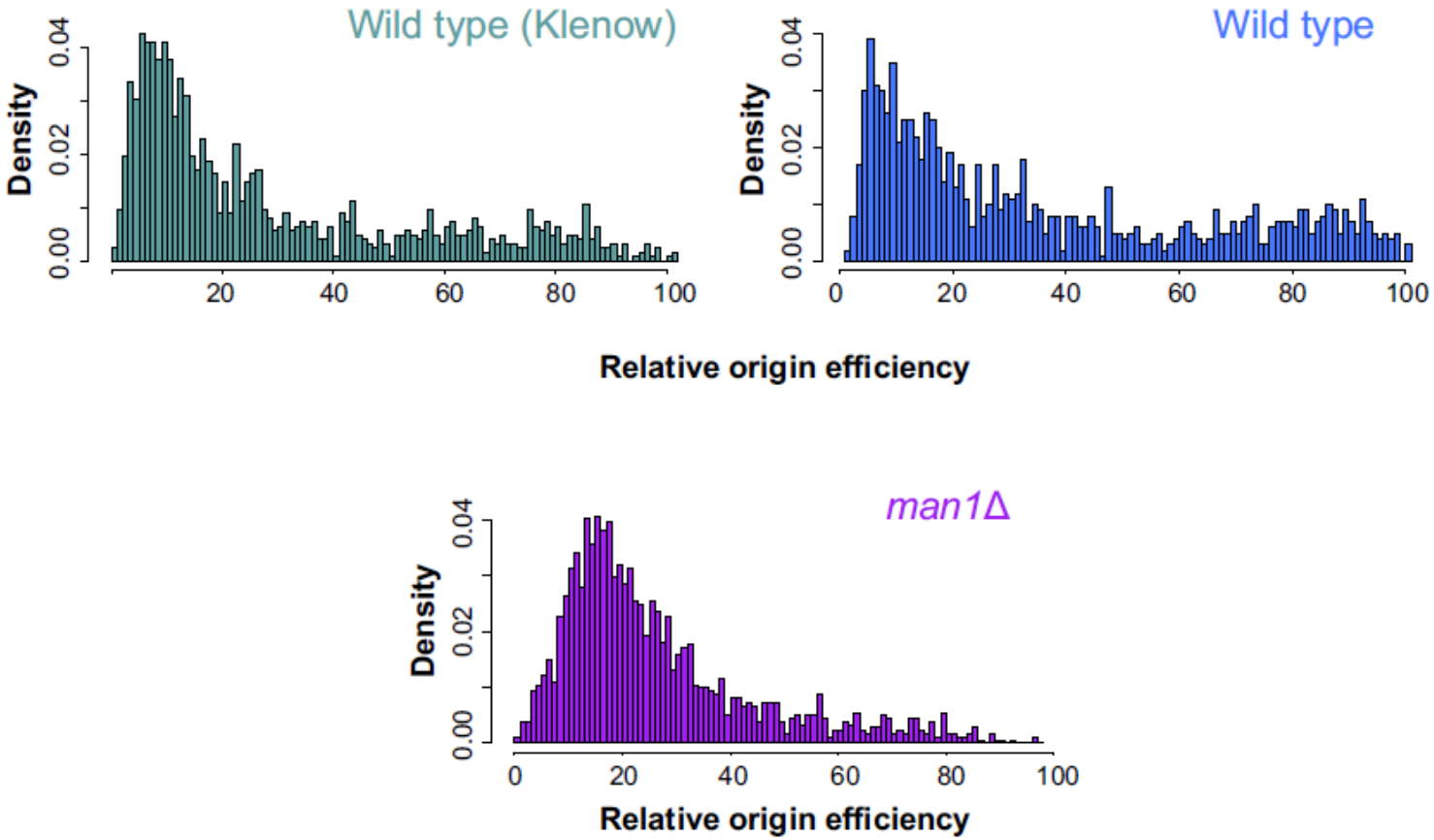
A) Polymerase usage across chromosome 3 in *man1*Δ and wild type *S. pombe*. Reads from all libraries were trimmed (30 bp at the 3' end and 1 bp at the 5' end) and aligned to the SP2 reference genome. The reads were binned and the number of reads in each bin was normalised to the total number of reads. Polymerase usage was calculated by taking the ratio of the normalised counts contributed by each polymerase to the total number of normalised counts contributed by both polymerases. The usage of polymerases δ and ε (shown in blue and red, respectively) on the forward and reverse strands were visualised using the Integrative Genome Viewer (IGV) genome browser. The wild type data shown here were generated from libraries that were constructed and sequenced alongside *man1*Δ libraries.

B) Pu-Seq Trep in *man1*Δ and wild type *S. pombe.* Pu-Seq Trep was calculated using the progression of leftward moving forks and assuming a constant fork velocity of 1.5 kb/min. It represents when in S-phase each locus is replicated. The y-axis was, therefore, inverted to maintain the convention of early and late replicating regions being shown on the top and bottom of the graph, respectively. The centromere is marked in grey. The wild type Pu-Seq Trep profile is derived from an average of data generated by 5 wild type Pu-Seq experiments.

The *rif1*Δ and wild type Pu-Seq Trep profiles are identical to those in Fig. 4.12 and are shown here only for comparison.

C) Pu-Seq Trep in *man1*Δ and wild type Klenow *S. pombe.* Pu-Seq Trep was calculated as described in B. The wild type Pu-Seq Trep profile was derived from a wild type Klenow experiment.

A) Density distributions of origin firing efficiencies



B) *man1Δ* origin activity genome wide

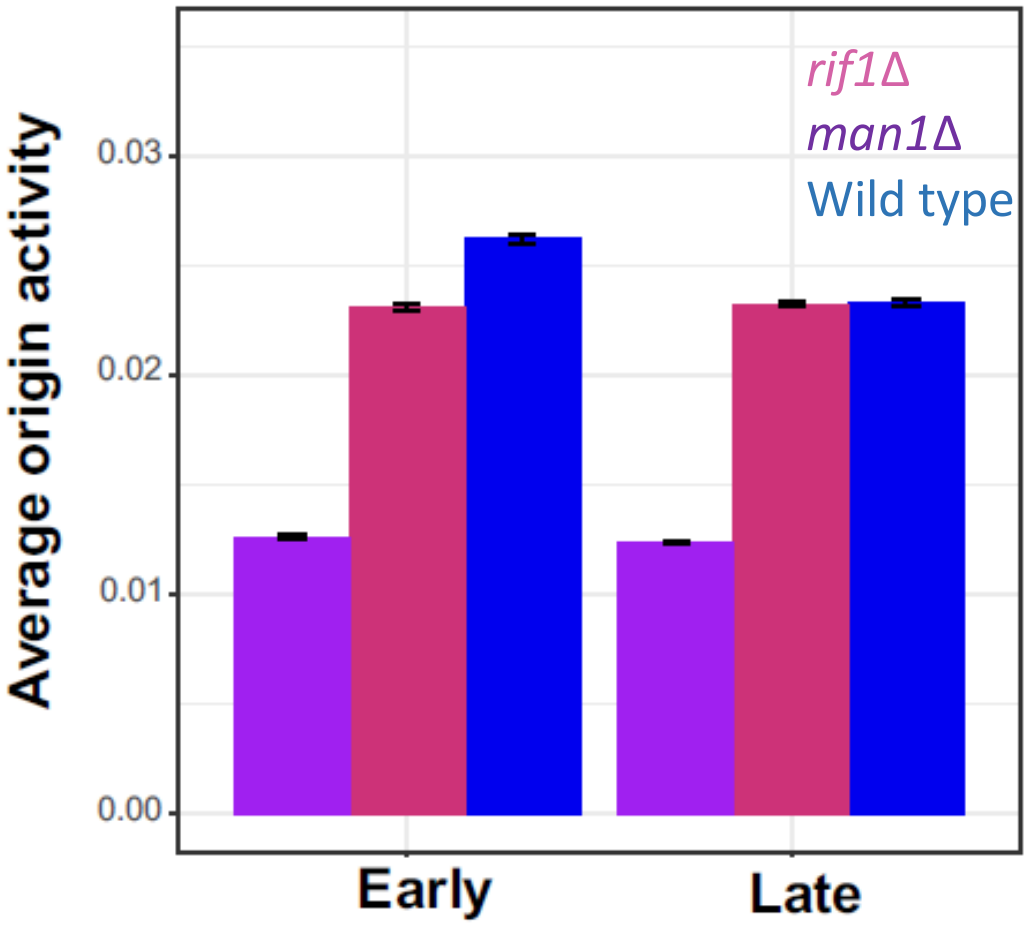


Fig. 6.4 **– Origin activity in *man1Δ* and wild type *S. pombe***

A) Density distribution of origin firing efficiencies in wild type and *man1Δ S. pombe*. Origins of replication were mapped for each strains using Pu-Seq, applying a threshold of 0.3, i.e., positive peaks of the differentials of polymerase usage whose heights were above the 30th percentile were mapped as origins. The efficiency of firing was normalised to the value of the 99th percentile in each data set.

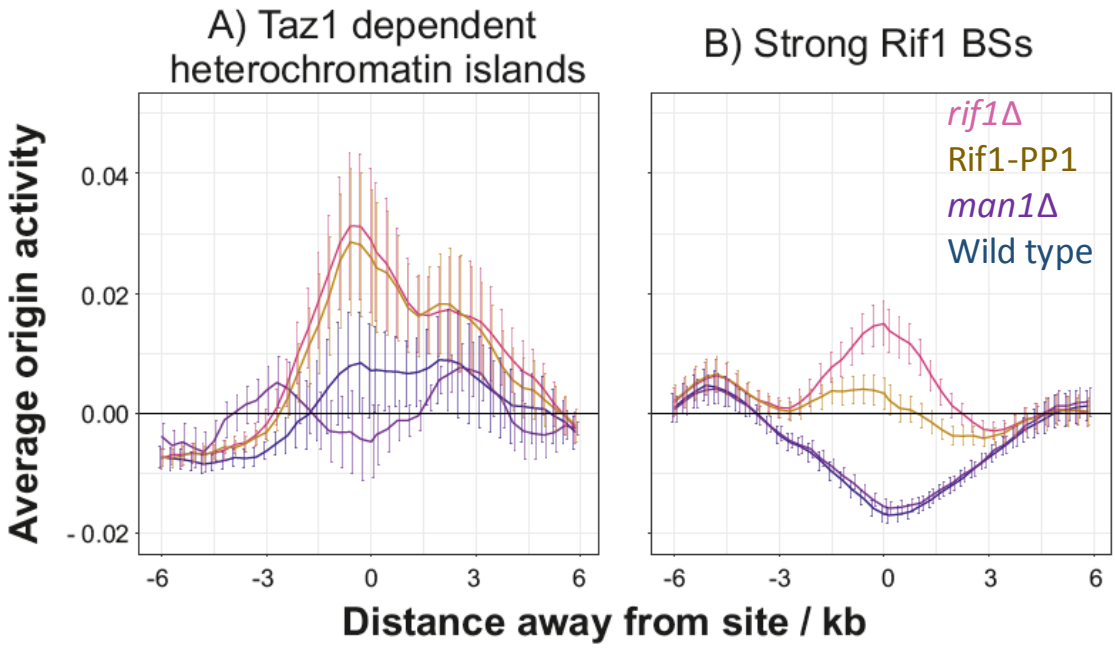
The wild type distribution (shown in dark bluw) is identical to that in Fig. 4.1 and is shown here only for comparison

B) The average origin across all late and early replicating regions in *rif1Δ*, *man1Δ* and wild type *S. pombe*. The average origin activity was calculated by averaging the differentials of the polymerase δ usage on the reverse strand and polymerase ε forward strand in all the bins spanning the late and early regions. Early replicating regions in *S. pombe* are marked in green on Fig. 4.9. The wild type and *rif1*Δ data are identical to those in Fig. 5.10 B and are shown here only for comparison.

6.2.3 Impact of *man1Δ* on local origin activity and replication timing

To determine whether the binding of Rif1 to DNA, and the subsequent local inhibition of origin firing, were a consequence of the Man1 dependent chromatin distribution, we analysed the origin activity around Taz1 dependent heterochromatin islands and strong Rif1 binding sites in *man1*Δ.

The origin activity up and downstream of Taz1 dependent heterochromatin islands and strong Rif1 BSs is shown in Figs. 6.5 A and B. In *man1*Δ, the origin activity around these sites was comparably low to that of wild type. Around strong Rif1 BSs, the origin activity decreased more sharply than it did in wild type - origin activity ~1.5 kb around the sites was affected, compared to 3 kb in wild type (Fig. 6.5 B). The local RT around the Taz1 dependent heterochromatin islands and Rif1 BSs followed an approximately wild type pattern (Figs. 6.5 C and D). Although the average RT around Rif1 BSs in *man1*Δ was greater than that in wild type (Fig. 6.5 D), there was a wild type "dip" in local RT at the positions of the binding site. These data show that the recruitment of Rif1 to Taz1 dependent heterochromatin islands and other Rif1 BSs is mostly independent on their position in the nucleus. Assuming that Rif1 localises to the nuclear periphery and associates with DNA in a Man1 independent manner, it is possible that Rif1 may play a minor role in the tethering of chromatin to the nuclear periphery.


A) Taz1 dependent heterochromatin islands
B) Strong Rif1 BSs
Average origin activity
0.04
0.02
0.00
- 0.02
-6
-3
0
3
6
rif1Δ
Rif1-PP1
man1Δ
Wild type
Distance away from site / kb

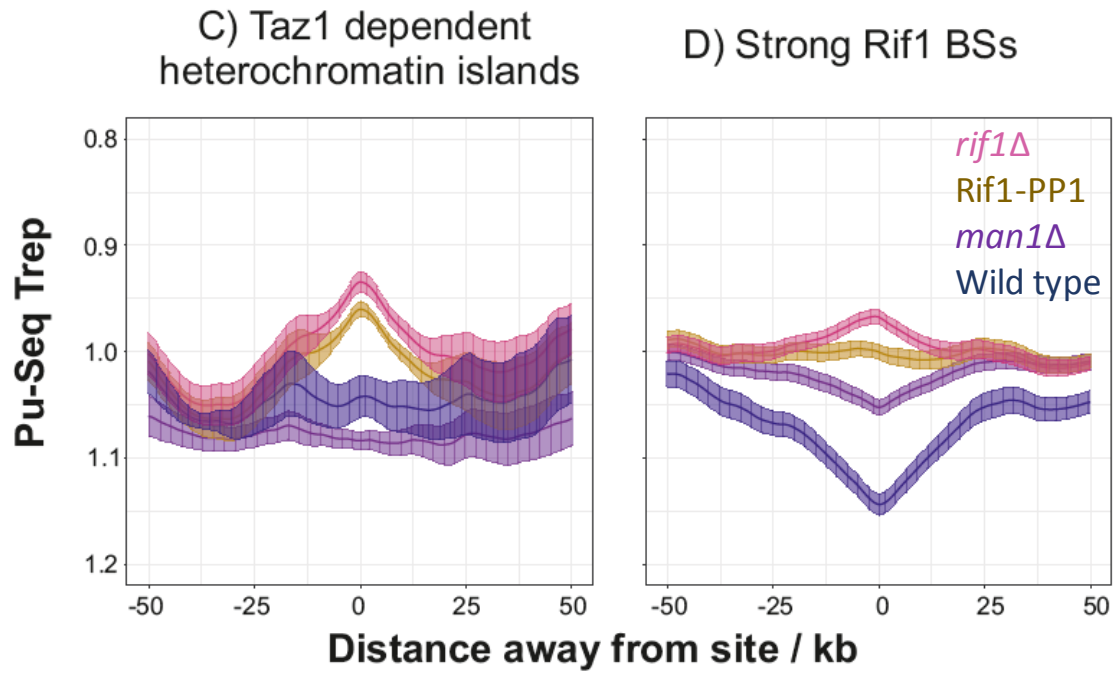

C) Taz1 dependent heterochromatin islands
D) Strong Rif1 BSs
Pu-Seq Trep
0.8
0.9
1.0
1.1
1.2
-50
-25
0
25
50
rif1Δ
Rif1-PP1
man1Δ
Wild type
Distance away from site / kb

Fig. 6.5 – **The average origin activity and Pu-Seq Trep around Taz1 dependent facultative heterochromatin islands and strong Rif1 BSs in *man1*Δ and wild type backgrounds**

The mean origin activity (i.e., the average of the differentials of the polymerase δ usage on the reverse strand and polymerase ε forward strand) in each 300 bp bin is shown for 6 kb up and downstream of the midpoint of each site.

Pu-Seq Trep was calculated using the progression of leftward moving forks and assuming a constant fork velocity of 1.5 kb/min. It represents when in S-phase at which each locus is replicated. The y-axis was, therefore inverted to maintain the convention of early and late replicating regions being shown on the top and bottom of the graph, respectively. Pu-Seq Trep 6 kb up and downstream of the midpoint of each site was recorded.

The error bars represent the standard error of the mean in each bin. The origin activity and Pu-Seq Trep shown here were calculated from two independent biological repeats for all backgrounds except *man1*Δ. The *rif1*Δ, Rif1-PP1 and wild type origin activity and Pu-Seq Trep data are identical to those shown in Fig. 5.6 and are only shown here for comparison.

A) The mean activity in each bin averaged across the 6 Taz1 dependent heterochromatin islands.

B) The mean activity in each bin averaged across the 35 strong Rif1 binding sites.

C) Pu-Seq Trep in each bin averaged across the 6 Taz1 dependent heterochromatin islands.

D) Pu-Seq Trep in each bin averaged across the 35 strong Rif1 binding sites.

6.2.4 Impact of heterochromatin on global RT program

As discussed in 1.4.3, heterochromatin does not to drive the formation of TADs, and therefore, the global RT program in metazoans (Dixon, Selvaraj et al. 2012, Nora, Lajoie et al. 2012). In *S. pombe*, *clr4*Δ does not impact the global distribution of chromatin and only tenuous links between the timing of origin firing and heterochromatin have been reported (Mizuguchi, Fudenberg et al. 2014, Pichugina, Sugawara et al. 2016).

To determine whether heterochromatin impacts the global RT program, we analysed *swi6*Δ and *clr4*Δ Pu-Seq Trep RT profiles. We did not observe a loss of the global RT in *clr4*Δ or *swi6*Δ (Fig. 6.6), suggesting that heterochromatin does not drive the formation of the nuclear distribution of chromosomal regions. The RT profile was only slightly distorted in late replicating regions in *clr4*Δ. It has been reported that *clr4*Δ may have some effect on the relative timing of firing of some origins (Mizuguchi, Fudenberg et al. 2014). It is possible that the small deviations in the *clr4*Δ RT profile are a result of these changes.

clr4Δ

swi6Δ

Wild type

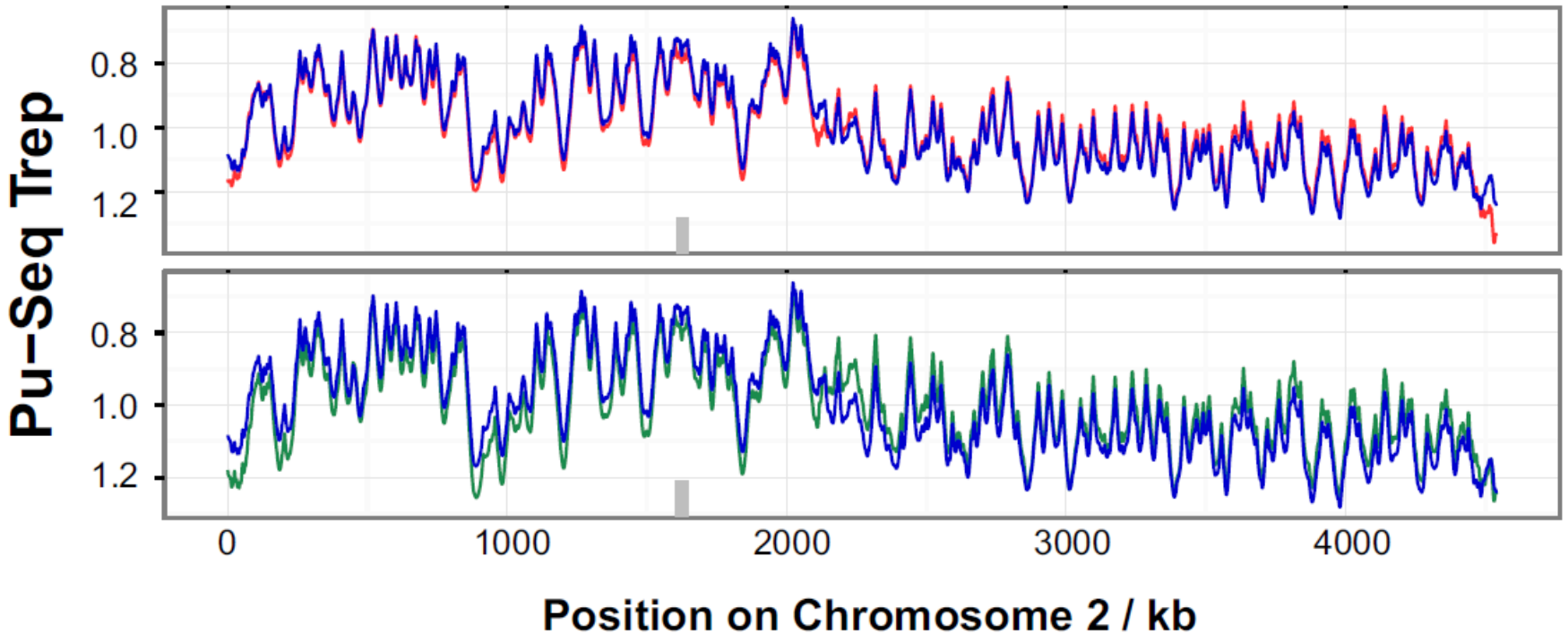


Fig. 6.6 - **Pu-Seq Trep across Chromosome 2 for *clr4*Δ, *swi6*Δ and wild type *S. pombe***

Pu-Seq Trep was calculated using the progression of leftward moving forks and assuming a constant fork velocity of 1.5 kb/min. It represents when in S-phase each locus is replicated. The y-axis was, therefore, inverted to maintain the convention of early and late replicating regions being shown on the top and bottom of the graph, respectively. The centromere is marked in grey.

6.2.5 *mrc1Δ* Elutri-Seq

We were unable to carry out a Pu-Seq analysis on *mrc1Δ* as the combination of *mrc1Δ* with the Pu-Seq background (*rnh201Δ cdc20-M630F* and *rnh201Δ cdc6-L591G*) was lethal. To test whether *mrc1Δ* has an impact on the global RT program, we carried out *mrc1Δ* Elutri-Seq.

mrc1Δ cells were synchronised in G_2 using centrifugal elutriation. $2.5x10^7$ G_2 cells were harvested and resuspended in rich YE media to a final concentration of $2.5x10^5$ cells/mL. The synchronisation 20 minutes after elutriation is shown in Fig. 6.7 A. The synchrony with which the cells progressed through the cell cycle was measured by the counting percentage of cells in each phase at every time point (Fig. 6.7 B) and using FACS (Fig. 6.7 C).

The synchronisation of *mrc1Δ* cells was successful - over 98% of cells were in G_2 20 minutes after elutriation. The cells cycle, however, did not progress with wild type dynamics (Fig 6.7 B). More cells persisted in M-phase and S-phase over longer time periods (Fig. 6.7 B) compared to *rif1Δ* (Fig. 5.2 B) or wild type cells (Daigaku, Keszthelyi et al. 2015). Despite the cells being at different stages of the cell cycle, the cells at each time point were of a similar size (Fig. 6.7). These data suggest that the progression of *mrc1Δ* cells through the cell cycle is not strictly length-independent.

The peak of septation was seen 80 minutes after elutriation (69% of cells in S-phase and 29% of cells in G_2). The 80 minute and 20 minute time points were taken as S-phase and G_2 samples, respectively, for *mrc1Δ* Elutri-Seq. The DNA was extracted, sonicated and used to prepare Illumina libraries as described in 2.2.5.1.

Despite the large number of reads generated by the libraries, very few reads aligned to the SP2 reference genome (Table 6.2), suggesting that the DNA fragments amplified for the Illumina library were predominantly not of *S. pombe* origin. A relatively low alignment rate was previously noted for Rif1-PP1 Elutri-Seq libraries (Table 5.3). The resultant coverage for Rif1-PP1 Elutri-Seq was sufficient for analysis and we did not investigate the cause of the low alignment rate further at the time. Taken together with the *mrc1Δ* Elutri-Seq data, it is possible that the process of elutriation introduces a contaminant whose genomic DNA is preferentially amplified during Illumina library preparation.

Table 6.2 – **Number of reads mapped and total coverage for *mrc1*Δ Elutri-Seq**

The total number of mapped reads is the sum of R1 and R2 mates that aligned in pairs (concordantly and discordantly) and single reads in single end mode. To maximise the coverage, untrimmed 80 bp reads were aligned.

Length of reads/bp	Sample	Total number of reads	Overall alignment rate	Number of reads mapped to the reference genome	Coverage
80	S-phase (80 min)	14,167,649	2.00%	283,353	1.8
	G_2 (20 min)	11,132,707	10.07%	1,121,064	7.1

Despite the low coverage, the aligned data were processed using the standard Elutri-Seq pipeline, as described in 2.2.5.2. The distribution of the *mrc1*Δ Elutri-Seq data is shown in Fig. 6.7. Given the very low coverage, the data are difficult to compare to that of wild type, *rif1*Δ (Fig. 5.3) or Rif1-PP1 (Fig. 5.8 D). The right slope of the histogram does resemble the "bulge" of early replicating regions (Fig. 6.8) seen in wild type Elutri-Seq distributions (Fig. 5.3). The data, however, are very noisy and correspond to different values on the x-axis.

To better determine the impact of *mrc1*Δ on global RT program, we directly compared the timing data between the early and late replicating regions on Chromosome 2 (as shown in Fig. 4.9). To do that, we calculated a ratio between the mean Elutri-Seq timing across the early and the late replicating regions. A similar ratio was calculated for the normalised number of reads scored in each bin from the S-phase sample (Reads S-phase ^{norm}). The loss of the global RT program in *S. pombe* produces a flat RT profile across Chromosome 2 (Fig. 5.7), which would result in a ratio value of 1. Relative differences between the early and late replicating domains would result in a value > 1. Both ratios were calculate for wild type, *rif1*Δ, Rif1-PP1 and *mrc1*Δ Elutri-Seq data and are shown in Table 6.3.

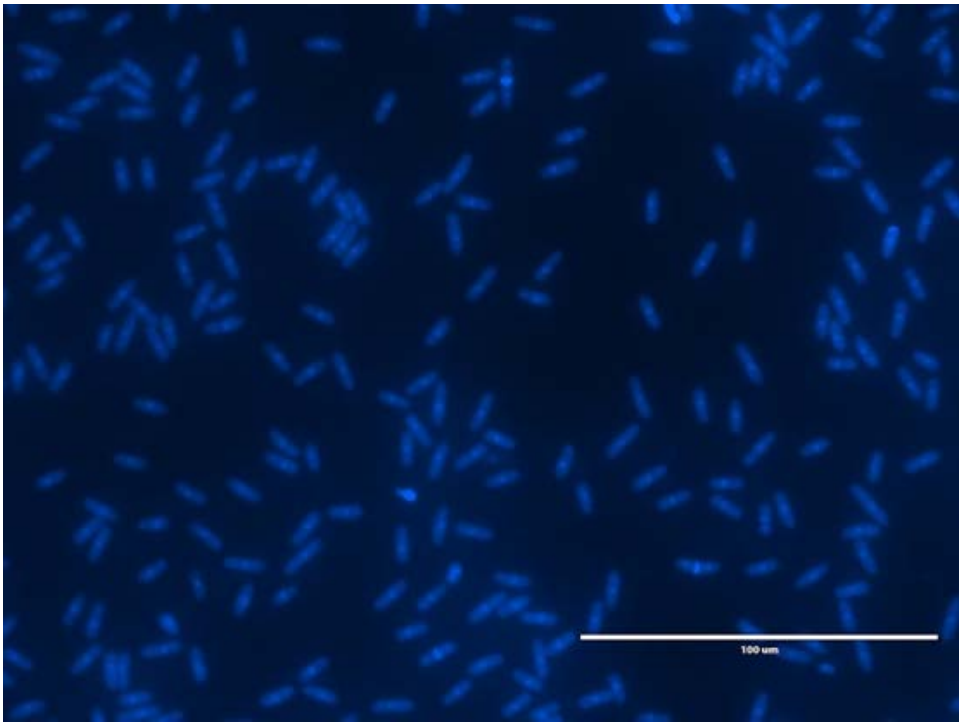
Table 6.3 - **Timing data ratios between early and late replicating regions on Chromosome 2**

Elutri-Seq ratios and ratios between the normalised number of reads scored in each bin from the S-phase sample were calculated from wild type, *rif1*Δ, Rif1-PP1 and *mrc1*Δ Elutri-Seq data

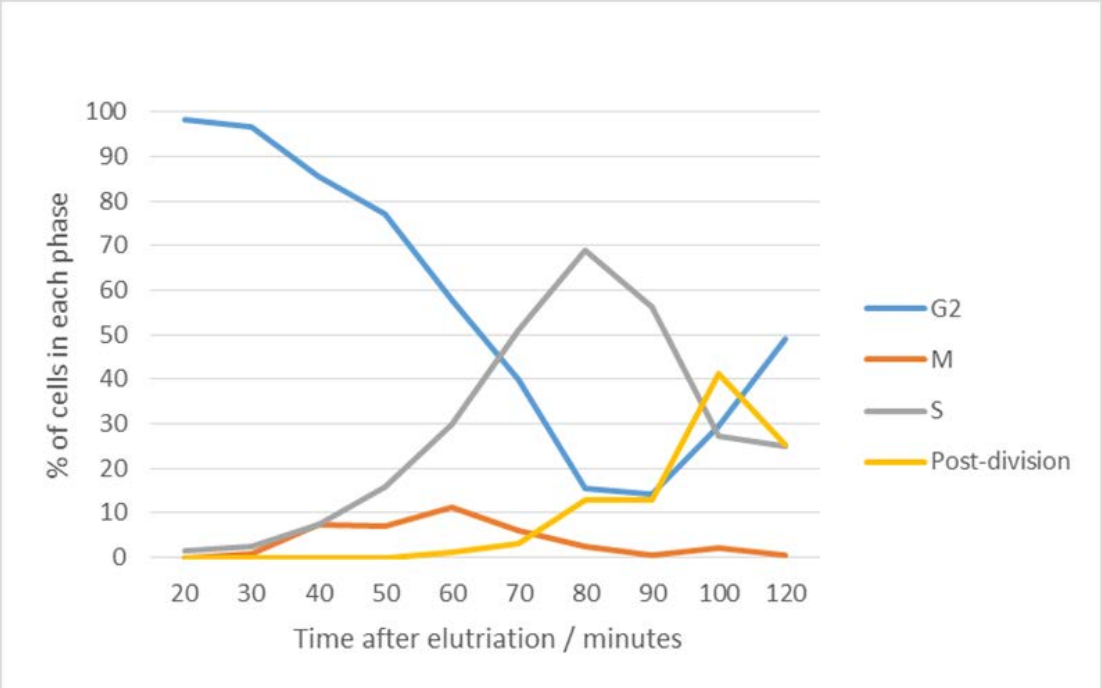
Elutri-Seq experiment	**Elutri-Seq $_{Early}$ / Elutri-Seq $_{Late}$**	**Reads S-phase $^{norm}{}_{Early}$ / Reads S-phase $^{norm}{}_{Late}$**
Wild type	1.1	1.1
***rif1*Δ**	1.0	1.0
Rif1-PP1	1.0	1.0
***mrc1*Δ**	1.2	1.1

The ratios between the early and late replicating regions suggest that the global RT program in *mrc1*Δ cells is wild type. It is, however, possible that the small proportion of G_2 cells in the *mrc1*Δ S-phase sample contributed to the increase in the number of reads across the early replicating regions.

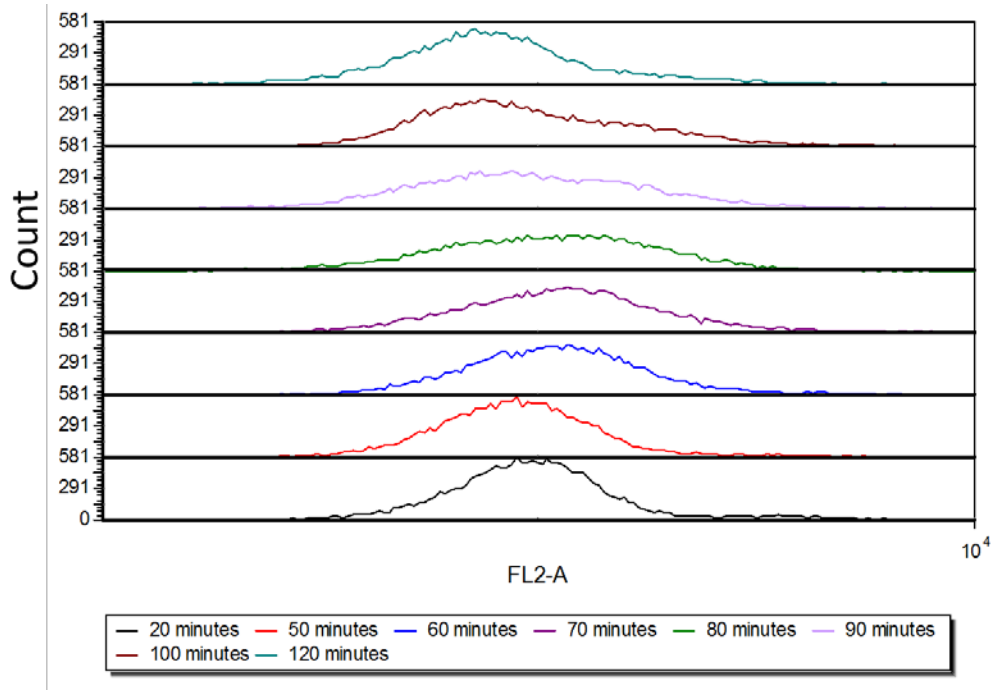
A) Synchronisation after 20 minutes



B) Septation index

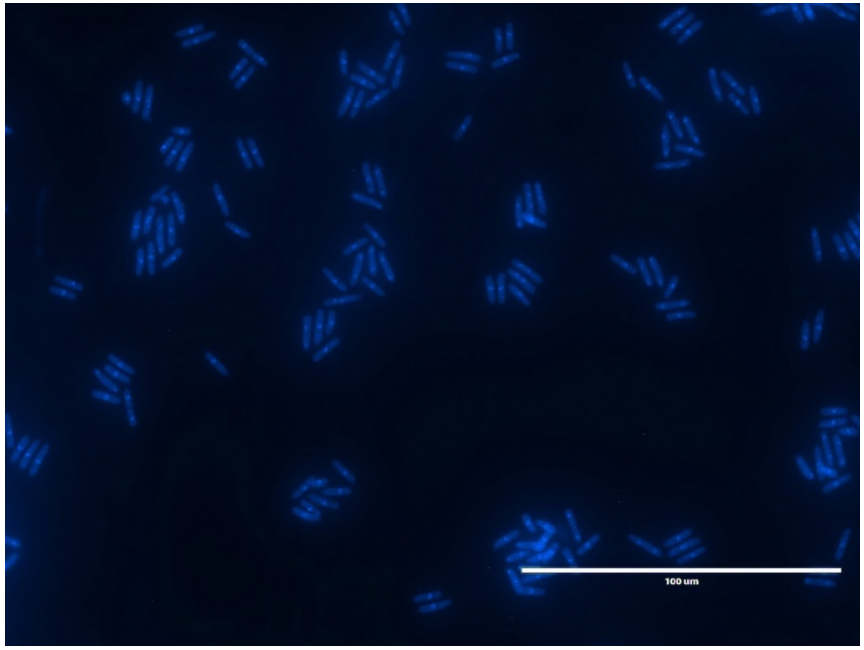


C) FACS


Count
581
291
581
291
581
291
581
291
581
291
581
291
581
291
581
291
0
10^4
FL2-A
20 minutes
50 minutes
60 minutes
70 minutes
80 minutes
90 minutes
100 minutes
120 minutes

D) Synchronisation after 20 and 80 minutes

50 minutes after elutriation



80 minutes after elutriation

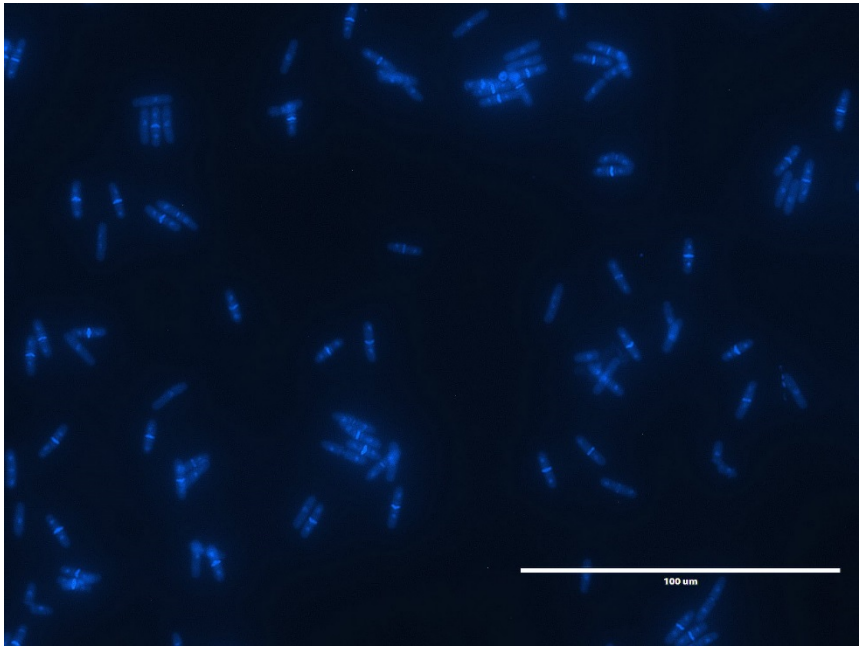


Fig. 6.7 - **Analysis of the cell cycle of *mrc1Δ* cells after elutriation**

A) Synchronisation of *mrc1Δ* 20 minutes after elutriation. Cells were stained with 1 µg/mL 4',6'-diamidino-2-phenylindole (DAPI) and 2.5% v/v calcofluor-white (which stain nucleic acid and the septum, respectively). Cells were visualized using an inverted fluorescence microscope (EVOS™ FL).

B) The synchronous passage of Rif1-PP1 cells through the cell cycle after elutriation. Cells were stained and visualised at each time point (as described in A). Cells were counted and divided into different phases of the cell cycle based on their morphology - cells with one nucleus - G_2 ; cells with two nuclei - M-phase ; cells with a septum- S-phase ; two cells joined, without a visible septum, and with one nucleus each - post-division.

C) FACS analysis of cells at each time point. $1.25x10^6$ cells were collected at each time point, stained with propidium iodide and analysed for DNA content on BD Accuri™ C6 Plus flow cytometer.

D) Synchronisation of *mrc1Δ* 50 and 80 minutes after elutriation. Cells were stained with 1 µg/mL 4',6'-diamidino-2-phenylindole (DAPI) and 2.5% v/v calcofluor-white (which stain nucleic acid and the septum, respectively). Cells were visualized using an inverted fluorescence microscope (EVOS™ FL).

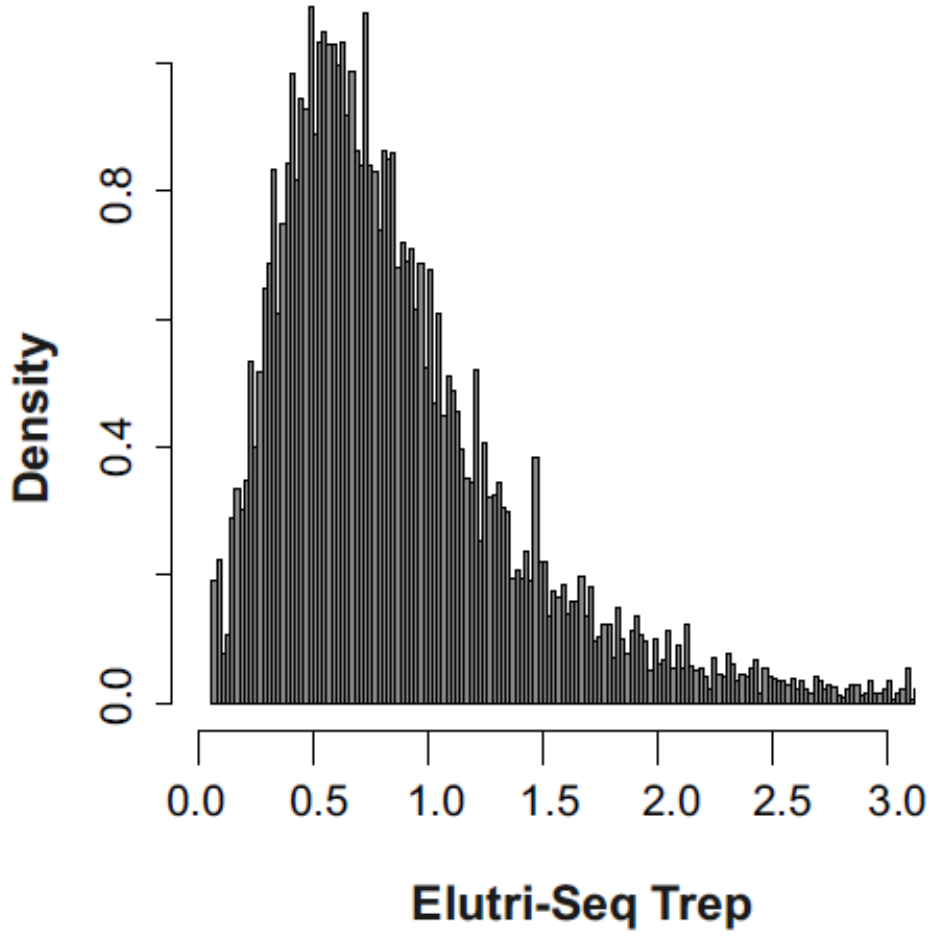


Fig. 6.8 - **Density distribution of Elutri-Seq Trep for *mrc1Δ S. pombe***

Density distribution of Elutri-Seq Trep for *mrc1*Δ1 *S. pombe*. Reads from S-phase and G_2 libraries (sequenced in for a coverage over 1 X) were aligned to the reference SP2 reference genome, binned into 1 kb windows, the number of 5' ends of reads was counted in each bin and normalised to the total number of counts. The ratio the S-phase and G_2 counts in every bin was calculated to produce Elutri-Seq Trep.

6.3 Discussion

6.3.1 Effect of nuclear distribution of chromatin on global RT program

Our data suggest that the Man1 dependent tethering of chromatin to the nuclear periphery is necessary, but not sufficient, to establish the global RT program in *S. pombe*. The data also hint that nuclear distribution of chromatin may regulate the global origin firing landscape.

When considering the consequences of *man1*Δ on origin firing and the global RT program, it is important to note that although the *man1*Δ polymerase usage ratios followed the same pattern as wild type, they were not as pronounced (Fig. 6.3 A). The narrow range of polymerase usage was similar to that described from the lower quality Klenow libraries (discussed in Chapter 3; Fig. 3.8). Unlike for the Klenow data, however, the genome coverage from the polymerase ε and δ libraries was very comparable (Table 6.1). While the Klenow wild type Pu-Seq Trep was the different to that calculated from T4 wild type data (Fig. 6.3 C), these variations did not explain all of the differences between wild type and *man1*Δ Pu-Seq Trep profiles. Additionally, more origins were mapped in *man1*Δ compared to wild type and the distribution of their normalised efficiencies was not bimodal (Fig. 6.4 A), with the majority of the origins having fired inefficiently (10% to 40% firing efficiency) in *man1*Δ It is, therefore, possible that the narrow range of polymerase usage is caused by a biological phenomenon in *man1*Δ. We cannot, however, exclude the possibility of a technical error without an independent biological repeat.

We propose that the tethering of chromatin to the nuclear periphery may regulate the landscape of origin firing. When the anchoring of chromatin is affect, the new nuclear positions of a subset of loci allow more regions to act as origins of replication. Given that origins fire stochastically, a large number of the new origins would fire inefficiently, explaining the distribution of global firing efficiencies (Fig. 6.4 A).

Regardless of the source, the narrow range of polymerase usage ratios in *man1*Δ resulted in a flatter global RT profile (Figs. 6.3 B and C) and lower absolute values of origin activity (Fig. 6.4 B). In comparison with a wild type Klenow Pu-Seq Trep RT profile (Fig. 6.3 C), the relative replication timing of subtelomeres and a ~1 Mb

region on the right arm of Chromosome 1 were affected in *man1*Δ. The delay in the regional RT on the arm of Chromosome 1 was most likely related to a decrease in the average origin firing activity in that region. Due to the global differences in the absolute values of origin activities between *man1*Δ and wild type, a meaningful comparison of the origin activity in that region was not possible.

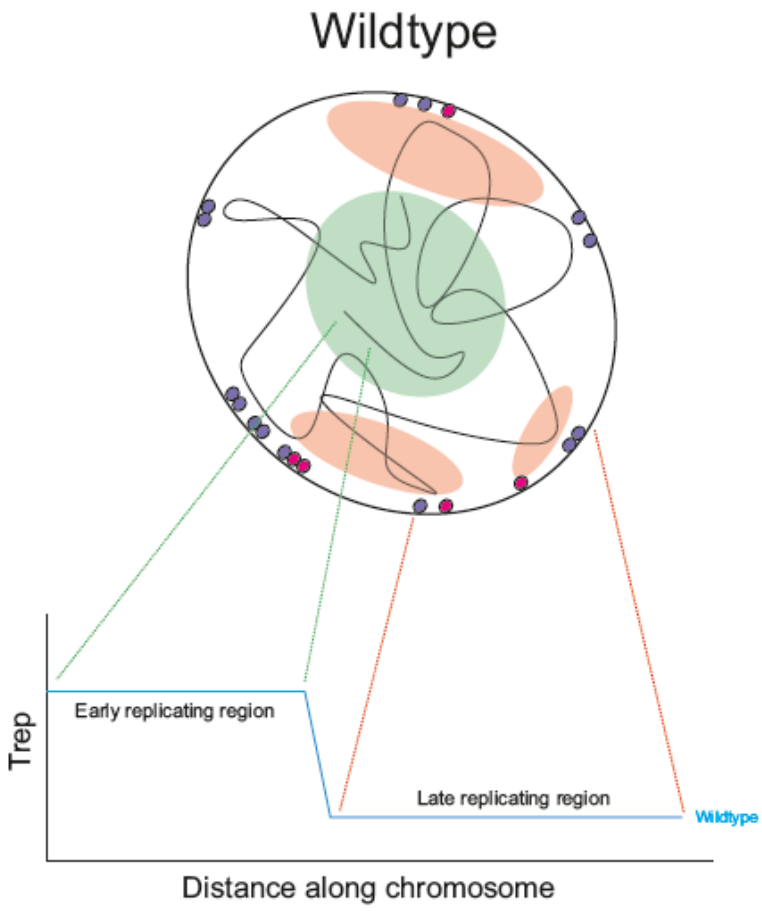
Interestingly, subtelomeres replicated early in *man1*Δ (Fig. 6.3 C), a phenotype reminiscent of that described for *taz1*Δ (Fig. 4.10 A) and *rif1*Δ (Fig. 4.12). The relative increase in the regional RT of the subtelomeres in *rif1*Δ and *taz1*Δ did not correlate with an increase in the origin activity in those regions (Fig. 4.7). In Chapter 5 we suggested that the change in RT can be explained by a model of replication where Rif1 mediated de-phosphorylation acts to set up late and early replicating regions of the genome, in addition to the local effects on origin inhibition. Based on the *man1*Δ Pu-Seq RT profile (Figs. 6.3 B and C), we propose that the Rif1 mediated inhibition of origin firing around subtelomeres may be driven by the Man1 dependent tethering of telomeres to the nuclear periphery.

We speculate that loss of chromatin tethering in *man1*Δ would result in different areas of the chromatin interacting with the nuclear periphery, which would, in turn, affect the regions acted on by Rif1 mediated PP1 phosphatases (i.e., regions "marked" for late replication). This would result in a global RT profile where only the parts of the genome not tethered to the periphery by Man1 would retain the late replicating phenotype in *man1*Δ (Fig. 6.9). The binding profiles of subtelomeric regions have been reported to be enriched in Man1 interactions (Steglich, Filion et al. 2012). Assuming that the loss of these interactions leads to changes in RT, this enrichment could account for a lot of the RT changes seen in *man1*Δ (Figs. 6.3 B and C).

Our data also show, however, that binding of Rif1 to its genomic binding sites (strong Rif1 BSs and Taz1 dependent heterochromatin islands) is not a consequence of the Man1 dependent nuclear distribution of chromatin (Fig. 6.5). Assuming that in S-phase most of Rif1 is tethered to the nuclear periphery, Rif1 may play a minor role in tethering chromatin to the nucleus. The effect of *rif1*Δ on global RT program, however, is unlikely to be caused by the loss of this tethering.

Our attempt to determine the localisation of endogenously expressed Rif1 using PALM was inconclusive. It is possible that the endogenous levels of Rif1 are lower than previously reported (Marguerat, Schmidt et al. 2012). It is also likely, however, that tagging Rif1 effects its expression and/or stability, resulting in very low levels of Rif1. The few nuclear localisation that were recorded, however, indicate that Rif1 (and Rif1-PP1) may localise to the nuclear and periphery (Fig. 6.2). It has been previously reported that the overexpression of Rif1-GFP results in a "nuclear haze" with distinct foci that co-localise with Taz1 (Zaaijer, Shaikh et al. 2016), indicating that Rif1 may exist in two populations - tethered to the nuclear periphery and freely diffusing in the nucleus. On the other hand, given its low endogenous levels of Rif1, it is unlikely that it exists in two separate pools and suggest that the reported "haze" may be an artefact of overexpression.

Despite not affecting the global distribution of chromatin in *S. pombe,* changes in the nuclear distribution of some origins of replication have been reported for *clr4*Δ (Mizuguchi, Fudenberg et al. 2014, Pichugina, Sugawara et al. 2016). Our data show no change in the global RT of *clr4*Δ and *swi6*Δ (Fig. 6.6). The RT profiles for *clr4*Δ did, however, show some small deviations from wild type. It is not clear whether these differences could be explained by experimental noise or whether they could be the consequence of the repositioning of some origins of replication. Regardless, the *clr4*Δ and *swi6*Δ global RT profiles support the idea that in *S. pombe*, similarly to metazoans (Dixon, Selvaraj et al. 2012, Nora, Lajoie et al. 2012), the nuclear distribution of chromatin is not driven by heterochromatin.


Wildtype
Trep
Early replicating region
Late replicating region
Wildtype
Distance along chromosome

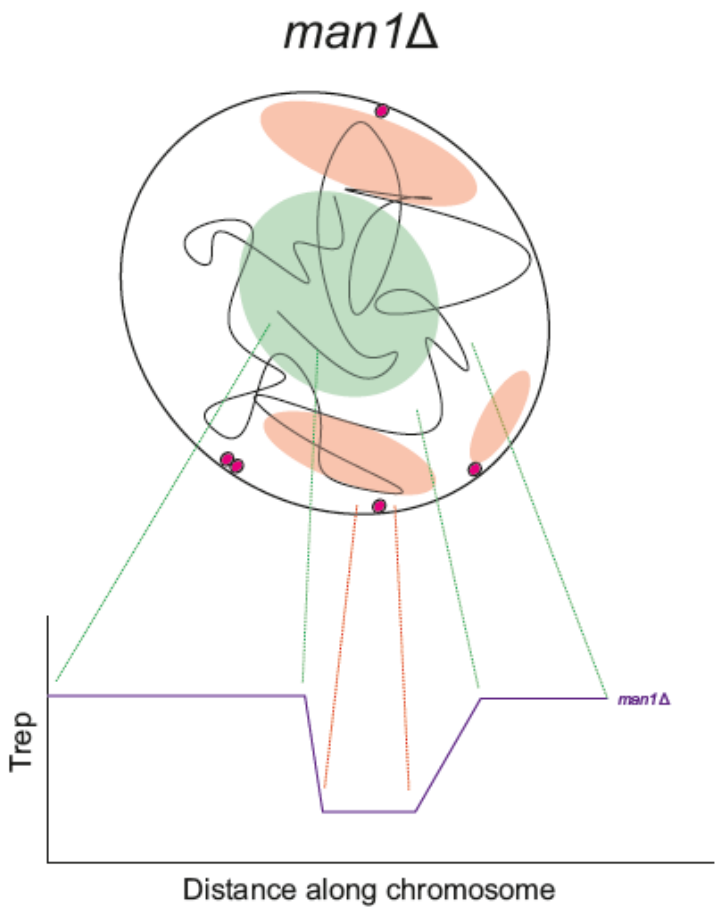

man1Δ
Trep
man1Δ
Distance along chromosome

Fig. 6.9 - **Proposed model of the effect of *man1*Δ on the global RT program**

In a wildtype background, chromatin is tethered to the nuclear periphery by Man1 (purple circles) and, to a smaller extent, by Rif1 (pink circles). The Rif1 mediated PP1 de-phosphorylation of Mcm2-7 is increased at the periphery, around Rif1. This de-phosphorylation establishes zones of late replication (shown in red) around the nuclear periphery. We propose that in these zones, replication is inhibited during the early S-phase (as discussed in Chapter 5). As a result, the origins in these regions will fire in late S-phase and the region will replicated later. Origins found on DNA located closer to the centre of the nucleus (i.e., in an early replicating zone - marked in green) fire in early S-phase. The regions in the red and green zones would correspond to late and early replicating regions on an RT profile, as indicated.

In *man1*Δ the tethering of chromatin to the nuclear periphery is affected. Rif1 mediated de-phosphorylation of Mcm2-7 still creates zones of early and late replicating regions. Different chromosomal regions, however, are present in these zones in *man1*Δ. In some cases, Rif1 can bind DNA (directly or via Taz1) and tether it to the nuclear periphery. These regions remain late replicating in *man1*Δ. In cases where Man1 acted to tether chromatin to late replicating zones created by Rif1, the late replicating phenotype is lost.

6.3.1 Role of Mrc1 in the global RT program of *S. pombe*

The *mrc1*Δ Elutri-Seq data are inconclusive. The synchronisation of the S-phase sample (Fig. 6.8) and the overall alignment rate of the reads from the S-phase and G_2 Elutri-Seq libraries were all sub-par. The low alignment rate of the *mrc1*Δ Elutri-Seq data (Table 6.2) was likely caused by a contaminant introduced to the *S. pombe* culture during elutriation. The contaminant was not visible when the cells were visualised and stained with DAPI (Figs. 5.2, 5.8 and 6.7). It is possible, however, that the low levels of genomic DNA introduced by the contaminant were preferentially amplified during Illumina library preparation.

The distribution of *mrc1*Δ Elutri-Seq (Fig. 6.8) shared some similarities with that of wild type *S. pombe* (Fig. 5.3). The data were not normally distributed - the bins on the right side of the x-axis formed a small hump, suggesting that some regions of the *mrc1*Δ genome replicate earlier than others. To determine whether this may be the case, we compared the mean Elutri-Seq and the mean number of reads in bins from the S-phase sample across early and late replicating regions on Chromosome 2 (Table 6.3). The ratios for both *rif1*Δ and Rif1-PP1, which exhibit a complete loss of the global RT program, were 1. On the other hand, ratios for *mrc1*Δ were > 1, which was similar to wild type, suggesting that the mean RT across the early and late regions of Chromosome 2 is not the same in *mrc1*Δ.

Given the presence of G_2 cells in the *mrc1*Δ S-phase sample (Fig. 6.8), the data need to be interpreted with caution. It is possible that the G_2 cells contributed to the increased number of reads across the early replicating regions and confounded the results. To conclusively determine the impact of *mrc1*Δ on the global RT program in *S. pombe*, the *mrc1*Δ Elutri-Seq RT profile needs to be repeated. While it is not necessary for standard elutriation, for future Elutri-Seq experiments we recommend using only sterilised equipment and working in a laminar flow cabinet. If the eradication of the contaminant is not possible, RT profiles can also be generated by standard marker frequency analysis using asynchronous cultures (method explained in 5.1.3).

Chapter 7

7.1 Conclusion

A number of proteins including, but no limited to LAP proteins Lem2 and Man1, act to tether chromatin to the nuclear periphery. We suggest that the overall nuclear distribution of chromatin driven by these proteins is necessary but sufficient to establish the genome wide RT program in *S. pombe*. Additionally, we propose that the nuclear distribution may be important for the regulation of the global origin firing landscape.

Our data also show that Rif1 may play a minor role in establishing some contacts between the chromatin and the nuclear periphery, either directly or via Taz1. One consequence of this is that Rif1 locally de-phosphorylates Mcm2-7 at a small number of origins that lie in regions adjacent to Rif1 binding sites. This inhibits origins from firing in the regions surrounding the Rif1 binding site, impacting the local RT around the site.

7.1 Maintenance of the global RT program by Rif1

The local effects of Rif1 described above are different to its role in the maintenance of the global RT program. The Rif1 mediated PP1 de-phosphorylation of Mcm2-7 in regions adjacent to Rif1 binding creates late replicating zones (shown in red on Fig. 7.1 Ai). We propose that in these zones, replication is inhibited in a PP1 dependant manner during early S-phase. At mid to late S-phase a signal (possibly a threshold phosphorylation of Rif1) is relayed to lift this inhibition. In our model we assume that most of Rif1 is bound to the nuclear periphery in G_1/S-phase. The late replicating zones are, therefore, concentrated around the nuclear periphery. Rif1, to a small extent, tethers some chromatin into the vicinity of the late replicating zones. Much of the chromatin found in the peripheral regions, and by extension in late replicating zones, is brought in by other proteins, e.g., Man1.

Origins in early replicating zones (shown in green on Fig. 7.1 Ai) are not inhibited from firing. In these zones, located towards the centre of the nucleus, the effective activity of CDK and DDK is greater than around the nuclear periphery (Fig. 7.1 Aii), as their targets are not being de-phosphorylated by Rif1 mediated PP1 phosphatases.

The establishment of the early and late replicating zones (Fig. 7.1 Ai), which are connected to the relative levels of Mcm2-7 phosphorylation (Fig. 7.1 Aii), results in the biphasic global RT program. During early S-phase, origins fire stochastically but only in the early replicating zones. In mid to late S-phase, when the Rif1 mediated de-phosphorylation is lifted, the effective concentration of kinase activity is increased around the nuclear periphery. Origin firing then proceeds with kinetics identical to those in early S-phase. This results in efficiently and inefficiently firing origins in both early and late replicating zones. Altogether, the global efficiency of origin firing in wild type *S. pombe* is distributed in a bimodal manner.

In *rif1*Δ and Rif1-PP1 cells, the distribution of chromatin around the nucleus is largely wild type. The early and late replicating zones, however, do not form given the lack of Rif1 and PP1 phosphatase interactions (Fig. 7.1 Bi) abolishing the gradient of effective kinase activity (Fig. 7.1 Bii). As a result of the interrupted interactions between Rif1 and PP1 phosphatases, Mcm2-7 hexamers at origins found in regions adjacent to the nuclear periphery are not de-phosphorylated. Consequently, no regions are preferentially replicated or inhibited from replicating during early S-phase. Instead, all origins have an equal probability of being replicated at any point (shown in orange in Fig. 7.1 Bi) in S-phase, resulting in a flat RT profile. Given that the distribution of chromatin in *rif1*Δ and Rif1-PP1 is largely unaffected, the global landscape of origin firing does not change, i.e., the same number of origins fire. Altogether, the global RT program is lost in *rif1*Δ and Rif1-PP1 but the global efficiencies of origin firing is distributed in a bimodal manner, similar to that of wild type.

7.2 Effects of Man1 and Mrc1 on global RT

In *man1*Δ, the Rif1 mediated early and late replication zones (Fig. 7.1 Ci) and the gradient of kinase activity (Fig. 7.1 Cii) are established in a wild type manner. Only the tethering of chromatin to the nuclear periphery is affected in *man1*Δ. Given, however, that not all of the interactions of chromatin with the nuclear periphery are driven by Man1, many of the same regions remain in close proximity to the nuclear periphery. This results in an intermediate loss of the global RT profile in *man1*Δ (Fig. 6.9). Origins in regions that are brought into the proximity of Rif1 in a Man1-independent manner are inhibited from firing during early S-phase and the

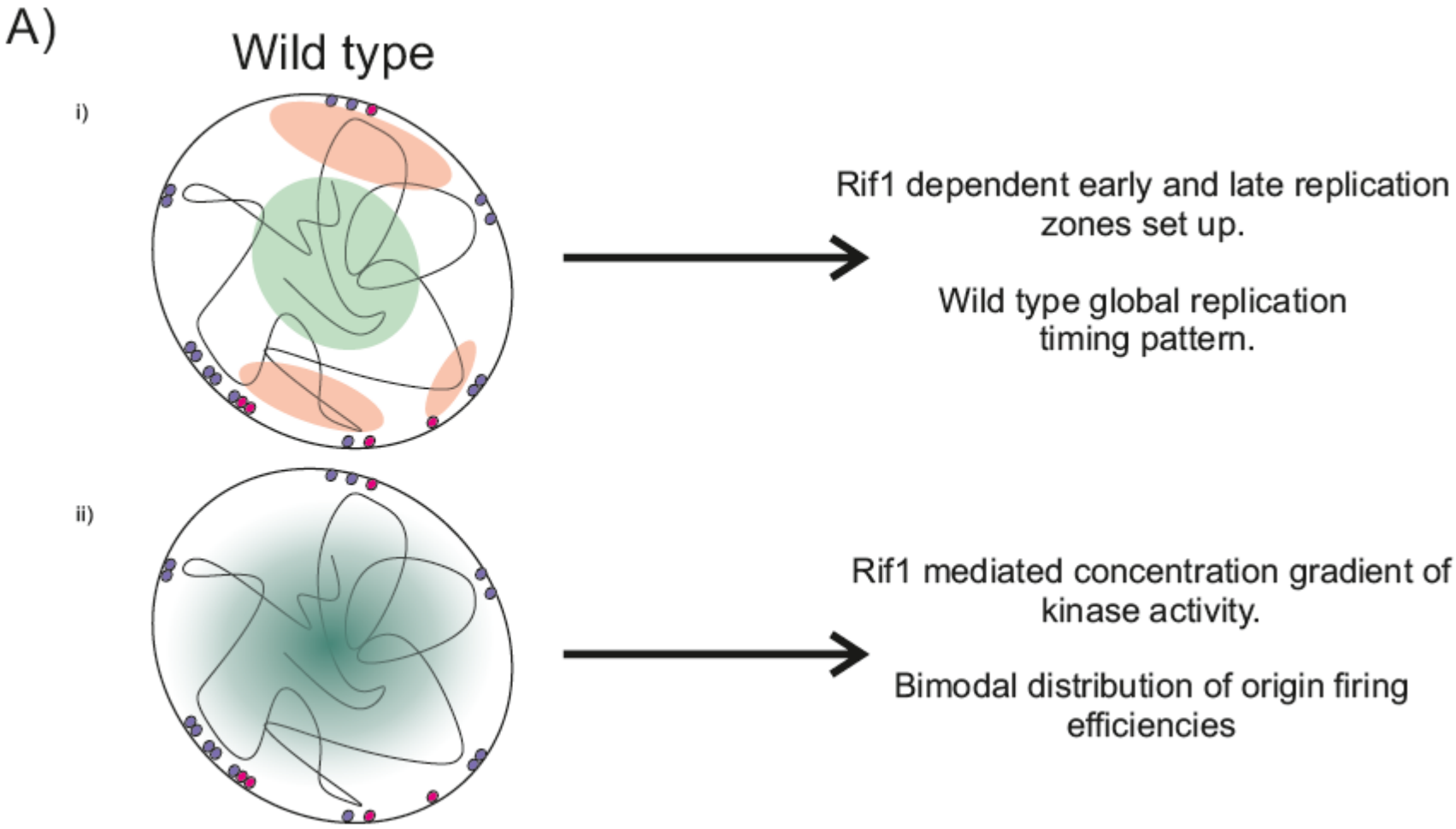
region maintains a late replicating phenotype in *man1*Δ. Regions tethered to the periphery by Man1 (e.g., subtelomeres) lose this association and can move into an early replicating zone which would affect their regional RT.

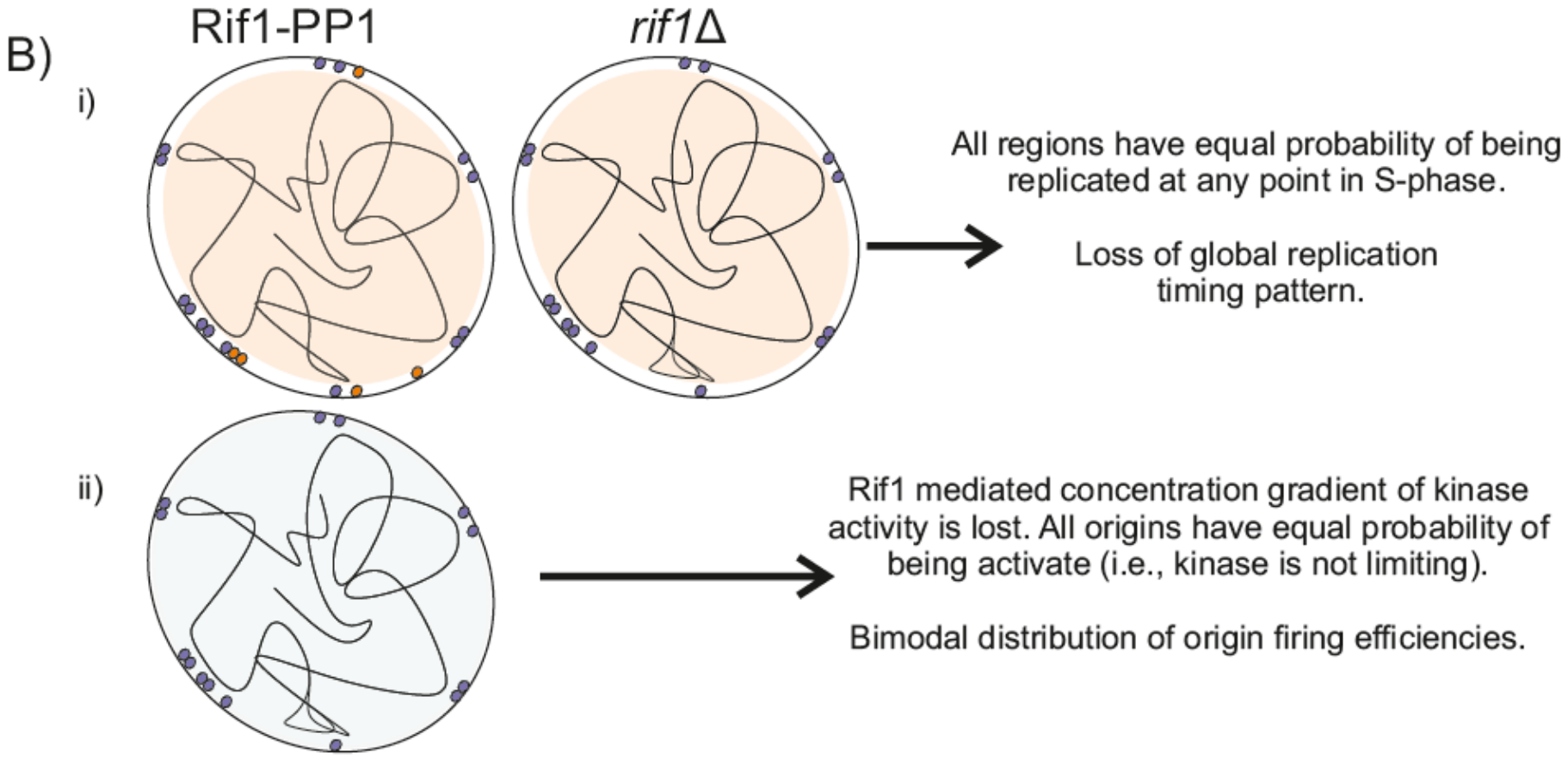
We propose that the increase in the number of origins fired seen in *man1*Δ was caused by the movement of chromosomal regions into areas of the nucleus that were more permissive to origin firing. In a wild type cells these loci would not have origin activity, possibly due to the local chromatin environment at the nuclear periphery. It is not clear what could lead to this inhibition. Given a wild type number of origins fired in *swi6*Δ and *clr4*Δ, it is unlikely that heterochromatin would play a role.

Alternatively, it is possible that the change in the global RT program in *man1*Δ is a consequence of the change in the landscape of origin firing. In this case, the change in the nuclear distribution of chromatin relative to a (Rif1 independent) limiting factor could impact the origin firing efficiencies and result in more inefficiently firing origins. The global change in origin firing could be a separate, Rif1-independent, way of modulating the global RT program.

This analysis and model do not address the role of cohesin dependent TAD-like globules in the global RT program or origin firing. Although no correlation between timing and globules has been reported, they do form and persist in G_1 (Mizuguchi, Fudenberg et al. 2014). The repositioning of chromosomal domains in G_1 is concomitant with the timing decision point (TDP) in metazoans (Dimitrova and Gilbert 1999). To further elucidate the role of chromatin on replication dynamics, the impact of cohesin loss on the maintenance of the global RT and origin firing should be studied.

Although our data are preliminary and need to be repeated, we did not find evidence that Mrc1 acts in parallel with Rif1 to establish the global RT program. We suggest that Mrc1 may act regulate the origins locally by recruiting Hsk1. It is also possible that the binding of Mrc1 to early firing origins in the early replicating zones could contribute to the increase of the effective kinase activity in the centre of the nucleus.

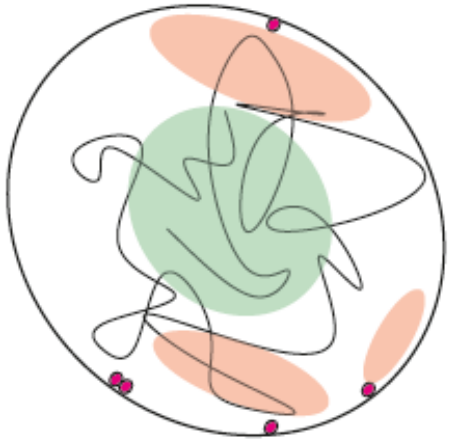

A)
Wild type
i)
Rif1 dependent early and late replication zones set up.
Wild type global replication timing pattern.
ii)
Rif1 mediated concentration gradient of kinase activity.
Bimodal distribution of origin firing efficiencies


B)
Rif1-PP1
rif1Δ
i)
All regions have equal probability of being replicated at any point in S-phase.
Loss of global replication timing pattern.
ii)
Rif1 mediated concentration gradient of kinase activity is lost. All origins have equal probability of being activate (i.e., kinase is not limiting).
Bimodal distribution of origin firing efficiencies.

C)

man1Δ

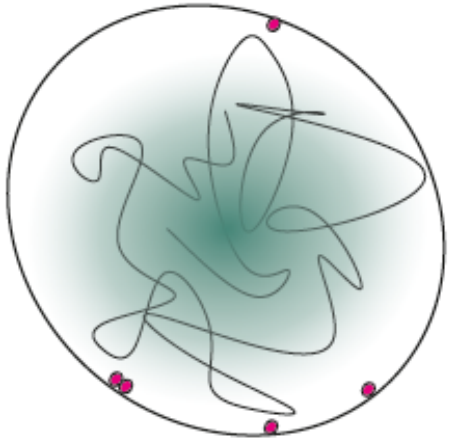
i)



Rif1 mediated early and late repliating zones set up. The distribution of chromatin within zones is different to wild type

Loss of late replicating phenotype in some regions

ii)



Total number of origins fired and the amount of DNA in the centre of the nucleus increases relative to wild type.

More origins fire inefficiently. Loss of bimodal distribution of origin firing efficiencies

Fig. 7.1 - **Proposed model of global replication timing and origin firing in *S. pombe***

A) In wild type cells (i) Man1 (purple circles) tethers chromatin to the nuclear periphery. Rif1 (pink circles) may contribute to the nuclear distribution of chromatin in a minor way. Rif1 interactions with phosphatases affect the relative distribution of kinase activity, creating zones of early and late DNA replication (marked in green and red, respectively). The effective kinase activity (shown in green in ii) is concentrated towards the centre of the nucleus. At the nuclear periphery, Rif1 mediated PP1 phosphatases dephosphorylate Mcm2-7, decreasing the effective kinase activity. This distribution in early S-phase inhibits origins in late zones from firing. When this inhibition is lifted in mid to late S-phase, the origins in late replicating zones fire stochastically. This allows efficient and inefficient firing of origins in late and early replicating regions and creates a biphasic global RT program.

B) When the interactions of Rif1 with the phosphatases are disrupted (Rif1-PP1) or in the absence of Rif1 (*rif1*Δ) the regions of early and late DNA replication disappear (i). The effective kinase activity is no longer limited to the peripheries in early S-phase (ii) and origins in all parts of the nucleus have an equal probability of firing at any point during S-phase. This results in a loss of the global biphasic RT profile. The global landscape of origin firing does not change, i.e., the global origin efficiencies are distributed in a bimodal manner.

C) In *man1*Δ the Rif1 dependent early and late replicating zones are established, similarly to wild type (i). The effective kinase activity is concentrated towards the centre of the nucleus and depleted around the periphery by PP1 phosphatases (ii). In *man1*Δ the nuclear distribution of chromatin is altered, affecting the loci located in the late and early replicating zones. This results in a biphasic global RT program but with different regions being replicated late and early, compared to wild type (see Fig. 6.9 for more detail). When not tethered to the periphery, more loci can act as origins of replication, increasing the total number of origins fired in *man1*Δ. Due to the stochastic nature of origin firing, most of the origins fire inefficiently.

8.0 Bibliography

Abe, T., K. Sugimura, Y. Hosono, Y. Takami, M. Akita, A. Yoshimura, S. Tada, T. Nakayama, H. Murofushi, K. Okumura, S. Takeda, M. Horikoshi, M. Seki and T. Enomoto (2011). "The histone chaperone facilitates chromatin transcription (FACT) protein maintains normal replication fork rates." J Biol Chem **286**(35): 30504-30512.

Alabert, C. and A. Groth (2012). "Chromatin replication and epigenome maintenance." Nat Rev Mol Cell Biol **13**(3): 153-167.

Alcasabas, A. A., A. J. Osborn, J. Bachant, F. Hu, P. J. Werler, K. Bousset, K. Furuya, J. F. Diffley, A. M. Carr and S. J. Elledge (2001). "Mrc1 transduces signals of DNA replication stress to activate Rad53." Nat Cell Biol **3**(11): 958-965.

Alvarez-Tabares, I., A. Grallert, J. M. Ortiz and I. M. Hagan (2007). "Schizosaccharomyces pombe protein phosphatase 1 in mitosis, endocytosis and a partnership with Wsh3/Tea4 to control polarised growth." J Cell Sci **120**(Pt 20): 3589-3601.

Alves-Rodrigues, I., P. G. Ferreira, A. Moldon, A. P. Vivancos, E. Hidalgo, R. Guigo and J. Ayte (2016). "Spatiotemporal Control of Forkhead Binding to DNA Regulates the Meiotic Gene Expression Program." Cell Rep **14**(4): 885-895.

Artem V. Artemov, Maria A. Andrianova, Georgii A. Bazykin and V. B. Seplyarskiy (2017). "POLD replicates both strands of small kilobase-long replication bubbles initiated at a majority of human replication origins." bioRxiv **174730**.

Ayoub, N., K. Noma, S. Isaac, T. Kahan, S. I. Grewal and A. Cohen (2003). "A novel jmjC domain protein modulates heterochromatization in fission yeast." Mol Cell Biol **23**(12): 4356-4370.

Bailis, J. M., P. Bernard, R. Antonelli, R. C. Allshire and S. L. Forsburg (2003). "Hsk1-Dfp1 is required for heterochromatin-mediated cohesion at centromeres." Nat Cell Biol **5**(12): 1111-1116.

Bakkenist, C. J. and M. B. Kastan (2004). "Initiating cellular stress responses." Cell **118**(1): 9-17.

Bando, M., Y. Katou, M. Komata, H. Tanaka, T. Itoh, T. Sutani and K. Shirahige (2009). "Csm3, Tof1, and Mrc1 form a heterotrimeric mediator complex that associates with DNA replication forks." J Biol Chem **284**(49): 34355-34365.

Baum, B., H. Nishitani, S. Yanow and P. Nurse (1998). "Cdc18 transcription and proteolysis couple S phase to passage through mitosis." EMBO J **17**(19): 5689-5698.

Bechhoefer, J. and N. Rhind (2012). "Replication timing and its emergence from stochastic processes." Trends Genet **28**(8): 374-381.

Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics **30**(15): 2114-2120.

Braun, S., J. F. Garcia, M. Rowley, M. Rougemaille, S. Shankar and H. D. Madhani (2011). "The Cul4-Ddb1(Cdt)(2) ubiquitin ligase inhibits invasion of a boundary-associated antisilencing factor into heterochromatin." Cell **144**(1): 41-54.

Breeden, L. L. (1997). "Alpha-factor synchronization of budding yeast." Methods Enzymol **283**: 332-341.

Brown, G. W. and T. J. Kelly (1999). "Cell cycle regulation of Dfp1, an activator of the Hsk1 protein kinase." Proc Natl Acad Sci U S A **96**(15): 8443-8448.

Bruck, I. and D. L. Kaplan (2011). "Origin single-stranded DNA releases Sld3 protein from the Mcm2-7 complex, allowing the GINS tetramer to bind the Mcm2-7 complex." J Biol Chem **286**(21): 18602-18613.

Bruck, I. and D. L. Kaplan (2014). "The replication initiation protein Sld2 regulates helicase assembly." J Biol Chem **289**(4): 1948-1959.

Buker, S. M., T. Iida, M. Buhler, J. Villen, S. P. Gygi, J. Nakayama and D. Moazed (2007). "Two different Argonaute complexes are required for siRNA generation and heterochromatin assembly in fission yeast." Nat Struct Mol Biol **14**(3): 200-207.

Bulmer, R., A. Pic-Taylor, S. K. Whitehall, K. A. Martin, J. B. Millar, J. Quinn and B. A. Morgan (2004). "The forkhead transcription factor Fkh2 regulates the cell division cycle of Schizosaccharomyces pombe." Eukaryot Cell **3**(4): 944-954.
Calzada, A., B. Hodgson, M. Kanemaki, A. Bueno and K. Labib (2005). "Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork." Genes Dev **19**(16): 1905-1919.
Cam, H. P., T. Sugiyama, E. S. Chen, X. Chen, P. C. FitzGerald and S. I. Grewal (2005). "Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome." Nat Genet **37**(8): 809-819.
Carr, A. M., A. L. Paek and T. Weinert (2011). "DNA replication: failures and inverted fusions." Semin Cell Dev Biol **22**(8): 866-874.
Carvajal-Yepes, M., L. Zambrano, J. M. Bueno, B. Raatz and W. J. Cuellar (2017). "Complete genome sequence of bean leaf crumple virus, a novel begomovirus infecting common bean in Colombia." Arch Virol **162**(6): 1773-1776.
Chatre, L., J. Fernandes, V. Michel, L. Fiette, P. Ave, G. Arena, U. Jain, R. Haas, T. C. Wang, M. Ricchetti and E. Touati (2017). "Helicobacter pylori targets mitochondrial import and components of mitochondrial DNA replication machinery through an alternative VacA-dependent and a VacA-independent mechanisms." Sci Rep **7**(1): 15901.
Chen, S., M. A. de Vries and S. P. Bell (2007). "Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2-7 loading." Genes Dev **21**(22): 2897-2907.
Chini, C. C. and J. Chen (2003). "Human claspin is required for replication checkpoint control." J Biol Chem **278**(32): 30057-30062.
Choi, E. S., Y. Cheon, K. Kang and D. Lee (2017). "The Ino80 complex mediates epigenetic centromere propagation via active removal of histone H3." Nat Commun **8**(1): 529.
Collins, K. A., A. R. Castillo, S. Y. Tatsutani and S. Biggins (2005). "De novo kinetochore assembly requires the centromeric histone H3 variant." Mol Biol Cell **16**(12): 5649-5660.
Conaway, R. C. and I. R. Lehman (1982). "A DNA primase activity associated with DNA polymerase alpha from Drosophila melanogaster embryos." Proc Natl Acad Sci U S A **79**(8): 2523-2527.
Cooley, C., A. Dave, M. Garg and A. Bianchi (2014). "Tel1ATM dictates the replication timing of short yeast telomeres." EMBO Rep **15**(10): 1093-1101.
Cornacchia, D., V. Dileep, J. P. Quivy, R. Foti, F. Tili, R. Santarella-Mellwig, C. Antony, G. Almouzni, D. M. Gilbert and S. B. Buonomo (2012). "Mouse Rif1 is a key regulator of the replication-timing programme in mammalian cells." EMBO J **31**(18): 3678-3690.
Coudreuse, D. and P. Nurse (2010). "Driving the cell cycle with a minimal CDK control network." Nature **468**(7327): 1074-1079.
Czajkowsky, D. M., J. Liu, J. L. Hamlin and Z. Shao (2008). "DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI." J Mol Biol **375**(1): 12-19.
Dai, J., R. Y. Chuang and T. J. Kelly (2005). "DNA replication origins in the Schizosaccharomyces pombe genome." Proc Natl Acad Sci U S A **102**(2): 337-342.
Daigaku, Y., A. Keszthelyi, C. A. Muller, I. Miyabe, T. Brooks, R. Retkute, M. Hubank, C. A. Nieduszyski and A. M. Carr (2015). "A global profile of replicative polymerase usage." Nat Struct Mol Biol **22**(3): 192-198.
Dalgaard, J. Z. and A. J. Klar (2001). "A DNA replication-arrest site RTS1 regulates imprinting by determining the direction of replication at mat1 in S. pombe." Genes Dev **15**(16): 2060-2068.
Dave, A., C. Cooley, M. Garg and A. Bianchi (2014). "Protein phosphatase 1 recruitment by Rif1 regulates DNA replication origin firing by counteracting DDK activity." Cell Rep **7**(1): 53-61.
De Piccoli, G., Y. Katou, T. Itoh, R. Nakato, K. Shirahige and K. Labib (2012). "Replisome stability at defective DNA replication forks is independent of S phase checkpoint kinases." Mol Cell **45**(5): 696-704.
Demeret, C., Y. Vassetzky and M. Mechali (2001). "Chromatin remodelling and DNA replication: from nucleosomes to loop domains." Oncogene **20**(24): 3086-3093.

Dhar, M. K., S. Sehgal and S. Kaul (2012). "Structure, replication efficiency and fragility of yeast ARS elements." Res Microbiol **163**(4): 243-253.
Dhingra, N., I. Bruck, S. Smith, B. Ning and D. L. Kaplan (2015). "Dpb11 protein helps control assembly of the Cdc45.Mcm2-7.GINS replication fork helicase." J Biol Chem **290**(12): 7586-7601.
Dileep, V., F. Ay, J. Sima, D. L. Vera, W. S. Noble and D. M. Gilbert (2015). "Topologically associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication-timing program." Genome Res **25**(8): 1104-1113.
Dimitrova, D. S. and D. M. Gilbert (1999). "The spatial position and replication timing of chromosomal domains are both established in early G1 phase." Mol Cell **4**(6): 983-993.
Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu and B. Ren (2012). "Topological domains in mammalian genomes identified by analysis of chromatin interactions." Nature **485**(7398): 376-380.
Donley, N. and M. J. Thayer (2013). "DNA replication timing, genome stability and cancer: late and/or delayed DNA replication timing is associated with increased genomic instability." Semin Cancer Biol **23**(2): 80-89.
Duan, Z., M. Andronescu, K. Schutz, S. McIlwain, Y. J. Kim, C. Lee, J. Shendure, S. Fields, C. A. Blau and W. S. Noble (2010). "A three-dimensional model of the yeast genome." Nature **465**(7296): 363-367.
Dungrawala, H., K. L. Rose, K. P. Bhat, K. N. Mohni, G. G. Glick, F. B. Couch and D. Cortez (2015). "The Replication Checkpoint Prevents Two Types of Fork Collapse without Regulating Replisome Stability." Mol Cell **59**(6): 998-1010.
Durkin, S. G. and T. W. Glover (2007). "Chromosome fragile sites." Annu Rev Genet **41**: 169-192.
Edwards, R. J., N. J. Bentley and A. M. Carr (1999). "A Rad3-Rad26 complex responds to DNA damage independently of other checkpoint proteins." Nat Cell Biol **1**(7): 393-398.
Ekwall, K., J. P. Javerzat, A. Lorentz, H. Schmidt, G. Cranston and R. Allshire (1995). "The chromodomain protein Swi6: a key component at fission yeast centromeres." Science **269**(5229): 1429-1431.
Ekwall, K., E. R. Nimmo, J. P. Javerzat, B. Borgstrom, R. Egel, G. Cranston and R. Allshire (1996). "Mutations in the fission yeast silencing factors clr4+ and rik1+ disrupt the localisation of the chromo domain protein Swi6p and impair centromere function." J Cell Sci **109 (Pt 11)**: 2637-2648.
Eser, U., D. Chandler-Brown, F. Ay, A. F. Straight, Z. Duan, W. S. Noble and J. M. Skotheim (2017). "Form and function of topologically associating genomic domains in budding yeast." Proc Natl Acad Sci U S A **114**(15): E3061-E3070.
Etheridge, T. J., R. L. Boulineau, A. Herbert, A. T. Watson, Y. Daigaku, J. Tucker, S. George, P. Jonsson, M. Palayret, D. Lando, E. Laue, M. A. Osborne, D. Klenerman, S. F. Lee and A. M. Carr (2014). "Quantification of DNA-associated proteins inside eukaryotic cells using single-molecule localization microscopy." Nucleic Acids Res **42**(19): e146.
Eydmann, T., E. Sommariva, T. Inagawa, S. Mian, A. J. Klar and J. Z. Dalgaard (2008). "Rtf1-mediated eukaryotic site-specific replication termination." Genetics **180**(1): 27-39.
Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver and C. C. Mello (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-811.
Folco, H. D., A. L. Pidoux, T. Urano and R. C. Allshire (2008). "Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres." Science **319**(5859): 94-97.
Foti, R., S. Gnan, D. Cornacchia, V. Dileep, A. Bulut-Karslioglu, S. Diehl, A. Buness, F. A. Klein, W. Huber, E. Johnstone, R. Loos, P. Bertone, D. M. Gilbert, T. Manke, T. Jenuwein and S. C. Buonomo (2015). "Nuclear Architecture Organized by Rif1 Underpins the Replication-Timing Program." Mol Cell.

Foti, R., S. Gnan, D. Cornacchia, V. Dileep, A. Bulut-Karslioglu, S. Diehl, A. Buness, F. A. Klein, W. Huber, E. Johnstone, R. Loos, P. Bertone, D. M. Gilbert, T. Manke, T. Jenuwein and S. C. Buonomo (2016). "Nuclear Architecture Organized by Rif1 Underpins the Replication-Timing Program." Mol Cell **61**(2): 260-273.
Gambus, A., R. C. Jones, A. Sanchez-Diaz, M. Kanemaki, F. van Deursen, R. D. Edmondson and K. Labib (2006). "GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks." Nat Cell Biol **8**(4): 358-366.
Garbacz, M. A., S. A. Lujan, A. B. Burkholder, P. B. Cox, Q. Wu, Z. X. Zhou, J. E. Haber and T. A. Kunkel (2018). "Evidence that DNA polymerase delta contributes to initiating leading strand DNA replication in Saccharomyces cerevisiae." Nat Commun **9**(1): 858.
Georgescu, R. E., L. Langston, N. Y. Yao, O. Yurieva, D. Zhang, J. Finkelstein, T. Agarwal and M. E. O'Donnell (2014). "Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork." Nat Struct Mol Biol **21**(8): 664-670.
Gilbert, D. M. (2002). "Replication timing and transcriptional control: beyond cause and effect." Curr Opin Cell Biol **14**(3): 377-383.
Glover, T. W., T. E. Wilson and M. F. Arlt (2017). "Fragile sites in cancer: more than meets the eye." Nature Reviews Cancer **17**(8): 489-501.
Gonzalez, Y., A. Saito and S. Sazer (2012). "Fission yeast Lem2 and Man1 perform fundamental functions of the animal cell nuclear lamina." Nucleus **3**(1): 60-76.
Gregan, J., K. Lindner, L. Brimage, R. Franklin, M. Namdar, E. A. Hart, S. J. Aves and S. E. Kearsey (2003). "Fission yeast Cdc23/Mcm10 functions after pre-replicative complex formation to promote Cdc45 chromatin binding." Mol Biol Cell **14**(9): 3876-3887.
Grewal, S. I. and S. C. Elgin (2007). "Transcription and RNA interference in the formation of heterochromatin." Nature **447**(7143): 399-406.
Grewal, S. I. and A. J. Klar (1997). "A recombinationally repressed region between mat2 and mat3 loci shares homology to centromeric repeats and regulates directionality of mating-type switching in fission yeast." Genetics **146**(4): 1221-1238.
Groth, A., A. Corpet, A. J. Cook, D. Roche, J. Bartek, J. Lukas and G. Almouzni (2007). "Regulation of replication fork progression through histone supply and demand." Science **318**(5858): 1928-1931.
Gruenbaum, Y. and R. Foisner (2015). "Lamins: nuclear intermediate filament proteins with fundamental functions in nuclear mechanics and genome regulation." Annu Rev Biochem **84**: 131-164.
Guarente, L. (2000). "Sir2 links chromatin silencing, metabolism, and aging." Genes Dev **14**(9): 1021-1026.
Guelen, L., L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, A. de Klein, L. Wessels, W. de Laat and B. van Steensel (2008). "Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions." Nature **453**(7197): 948-951.
Guilliam, T. A. and A. J. Doherty (2017). "PrimPol-Prime Time to Reprime." Genes (Basel) **8**(1).
Hafner, L., A. Lezaja, X. Zhang, L. Lemmens, M. Shyian, B. Albert, C. Follonier, J. M. Nunes, M. Lopes, D. Shore and S. Mattarocci (2018). "Rif1 Binding and Control of Chromosome-Internal DNA Replication Origins Is Limited by Telomere Sequestration." Cell Rep **23**(4): 983-992.
Hagan, I. M., A. Grallert and V. Simanis (2016). "Cell Cycle Synchronization of Schizosaccharomyces pombe by Centrifugal Elutriation of Small Cells." Cold Spring Harb Protoc **2016**(6): pdb prot091231.
Halic, M. and D. Moazed (2010). "Dicer-independent primal RNAs trigger RNAi and heterochromatin formation." Cell **140**(4): 504-516.
Hall, I. M., K. Noma and S. I. Grewal (2003). "RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast." Proc Natl Acad Sci U S A **100**(1): 193-198.

Handa, T., M. Kanke, T. S. Takahashi, T. Nakagawa and H. Masukata (2012). "DNA polymerization-independent functions of DNA polymerase epsilon in assembly and progression of the replisome in fission yeast." Mol Biol Cell **23**(16): 3240-3253.
Harigaya, Y., H. Tanaka, S. Yamanaka, K. Tanaka, Y. Watanabe, C. Tsutsumi, Y. Chikashige, Y. Hiraoka, A. Yamashita and M. Yamamoto (2006). "Selective elimination of messenger RNA prevents an incidence of untimely meiosis." Nature **442**(7098): 45-50.
Hayano, M., Y. Kanoh, S. Matsumoto and H. Masai (2011). "Mrc1 marks early-firing origins and coordinates timing and efficiency of initiation in fission yeast." Mol Cell Biol **31**(12): 2380-2391.
Hayano, M., Y. Kanoh, S. Matsumoto, C. Renard-Guillet, K. Shirahige and H. Masai (2012). "Rif1 is a global regulator of timing of replication origin firing in fission yeast." Genes Dev **26**(2): 137-150.
Hayashi, M. T., T. S. Takahashi, T. Nakagawa, J. Nakayama and H. Masukata (2009). "The heterochromatin protein Swi6/HP1 activates replication origins at the pericentromeric region and silent mating-type locus." Nat Cell Biol **11**(3): 357-362.
Hiraga, S., G. M. Alvino, F. Chang, H. Y. Lian, A. Sridhar, T. Kubota, B. J. Brewer, M. Weinreich, M. K. Raghuraman and A. D. Donaldson (2014). "Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse Cdc7-mediated phosphorylation of the MCM complex." Genes Dev **28**(4): 372-383.
Hiratani, I., T. Ryba, M. Itoh, J. Rathjen, M. Kulik, B. Papp, E. Fussner, D. P. Bazett-Jones, K. Plath, S. Dalton, P. D. Rathjen and D. M. Gilbert (2010). "Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis." Genome Res **20**(2): 155-169.
Hiratani, I., T. Ryba, M. Itoh, T. Yokochi, M. Schwaiger, C. W. Chang, Y. Lyou, T. M. Townes, D. Schubeler and D. M. Gilbert (2008). "Global reorganization of replication domains during embryonic stem cell differentiation." PLoS Biol **6**(10): e245.
Hofmann, J. F. and D. Beach (1994). "cdt1 is an essential target of the Cdc10/Sct1 transcription factor: requirement for DNA replication and inhibition of mitosis." EMBO J **13**(2): 425-434.
Hong, E. J., J. Villen, E. L. Gerace, S. P. Gygi and D. Moazed (2005). "A cullin E3 ubiquitin ligase complex associates with Rik1 and the Clr4 histone H3-K9 methyltransferase and is required for RNAi-mediated heterochromatin formation." RNA Biol **2**(3): 106-111.
Horn, P. J., J. N. Bastie and C. L. Peterson (2005). "A Rik1-associated, cullin-dependent E3 ubiquitin ligase is essential for heterochromatin formation." Genes Dev **19**(14): 1705-1714.
Hsieh, T. H., A. Weiner, B. Lajoie, J. Dekker, N. Friedman and O. J. Rando (2015). "Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C." Cell **162**(1): 108-119.
Ikegami, K., T. A. Egelhofer, S. Strome and J. D. Lieb (2010). "Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2." Genome Biol **11**(12): R120.
Ilves, I., T. Petojevic, J. J. Pesavento and M. R. Botchan (2010). "Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins." Mol Cell **37**(2): 247-258.
Iraqui, I., Y. Chekkal, N. Jmari, V. Pietrobon, K. Freon, A. Costes and S. A. Lambert (2012). "Recovery of arrested replication forks by homologous recombination is error-prone." PLoS Genet **8**(10): e1002976.
Irvine, D. V., M. Zaratiegui, N. H. Tolia, D. B. Goto, D. H. Chitwood, M. W. Vaughn, L. Joshua-Tor and R. A. Martienssen (2006). "Argonaute slicing is required for heterochromatic silencing and spreading." Science **313**(5790): 1134-1137.
Ivanova, A. V., M. J. Bonaduce, S. V. Ivanov and A. J. Klar (1998). "The chromo and SET domains of the Clr4 protein are essential for silencing in fission yeast." Nat Genet **19**(2): 192-195.
Jacob, F. and S. Brenner (1963). "[On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon]." C R Hebd Seances Acad Sci **256**: 298-300.
Jia, S., R. Kobayashi and S. I. Grewal (2005). "Ubiquitin ligase component Cul4 associates with Clr4 histone methyltransferase to assemble heterochromatin." Nat Cell Biol **7**(10): 1007-1013.
Jia, S., K. Noma and S. I. Grewal (2004). "RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins." Science **304**(5679): 1971-1976.

Jiang, W., D. McDonald, T. J. Hope and T. Hunter (1999). "Mammalian Cdc7-Dbf4 protein kinase complex is essential for initiation of DNA replication." EMBO J **18**(20): 5703-5713.
Kamimura, Y., Y. S. Tak, A. Sugino and H. Araki (2001). "Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in Saccharomyces cerevisiae." EMBO J **20**(8): 2097-2107.
Kang, Y. H., W. C. Galal, A. Farina, I. Tappin and J. Hurwitz (2012). "Properties of the human Cdc45/Mcm2-7/GINS helicase complex and its action with DNA polymerase epsilon in rolling circle DNA synthesis." Proc Natl Acad Sci U S A **109**(16): 6042-6047.
Kanoh, J. and F. Ishikawa (2001). "spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast." Curr Biol **11**(20): 1624-1630.
Kanoh, J., M. Sadaie, T. Urano and F. Ishikawa (2005). "Telomere binding protein Taz1 establishes Swi6 heterochromatin independently of RNAi at telomeres." Curr Biol **15**(20): 1808-1819.
Kanoh, Y., S. Matsumoto, R. Fukatsu, N. Kakusho, N. Kono, C. Renard-Guillet, K. Masuda, K. Iida, K. Nagasawa, K. Shirahige and H. Masai (2015). "Rif1 binds to G quadruplexes and suppresses replication over long distances." Nat Struct Mol Biol **22**(11): 889-897.
Katou, Y., Y. Kanoh, M. Bando, H. Noguchi, H. Tanaka, T. Ashikari, K. Sugimoto and K. Shirahige (2003). "S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex." Nature **424**(6952): 1078-1083.
Kaykov, A. and P. Nurse (2015). "The spatial and temporal organization of origin firing during the S-phase of fission yeast." Genome Res **25**(3): 391-401.
Kelly, T. J., G. S. Martin, S. L. Forsburg, R. J. Stephen, A. Russo and P. Nurse (1993). "The fission yeast cdc18+ gene product couples S phase to START and mitosis." Cell **74**(2): 371-382.
Keszthelyi, A., Y. Daigaku, K. Ptasinska, I. Miyabe and A. M. Carr (2015). "Mapping ribonucleotides in genomic DNA and exploring replication dynamics by polymerase usage sequencing (Pu-seq)." Nat Protoc **10**(11): 1786-1801.
Kim, S. M., D. D. Dubey and J. A. Huberman (2003). "Early-replicating heterochromatin." Genes Dev **17**(3): 330-335.
Knott, S. R., J. M. Peace, A. Z. Ostrow, Y. Gan, A. E. Rex, C. J. Viggiani, S. Tavare and O. M. Aparicio (2012). "Forkhead transcription factors establish origin timing and long-range clustering in S. cerevisiae." Cell **148**(1-2): 99-111.
Knutsen, J. H., I. D. Rein, C. Rothe, T. Stokke, B. Grallert and E. Boye (2011). "Cell-cycle analysis of fission yeast cells by flow cytometry." PLoS One **6**(2): e17175.
Kobayashi, T. and T. Horiuchi (1996). "A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities." Genes Cells **1**(5): 465-474.
Krings, G. and D. Bastia (2004). "swi1- and swi3-dependent and independent replication fork arrest at the ribosomal DNA of Schizosaccharomyces pombe." Proc Natl Acad Sci U S A **101**(39): 14085-14090.
Kuipers, M. A., T. J. Stasevich, T. Sasaki, K. A. Wilson, K. L. Hazelwood, J. G. McNally, M. W. Davidson and D. M. Gilbert (2011). "Highly stable loading of Mcm proteins onto chromatin in living cells requires replication to unload." J Cell Biol **192**(1): 29-41.
Lambert, S. and A. M. Carr (2005). "Checkpoint responses to replication fork barriers." Biochimie **87**(7): 591-602.
Lambert, S. and A. M. Carr (2013). "Impediments to replication fork movement: stabilisation, reactivation and genome instability." Chromosoma **122**(1-2): 33-45.
Lambert, S., S. J. Mason, L. J. Barber, J. A. Hartley, J. A. Pearce, A. M. Carr and P. J. McHugh (2003). "Schizosaccharomyces pombe checkpoint response to DNA interstrand cross-links." Mol Cell Biol **23**(13): 4728-4737.
Lambert, S., K. Mizuno, J. Blaisonneau, S. Martineau, R. Chanet, K. Freon, J. M. Murray, A. M. Carr and G. Baldacci (2010). "Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange." Mol Cell **39**(3): 346-359.

Lander, E. S. and M. S. Waterman (1988). "Genomic mapping by fingerprinting random clones: a mathematical analysis." Genomics **2**(3): 231-239.
Larasati and B. P. Duncker (2016). "Mechanisms Governing DDK Regulation of the Initiation of DNA Replication." Genes (Basel) **8**(1).
Lebofsky, R., R. Heilig, M. Sonnleitner, J. Weissenbach and A. Bensimon (2006). "DNA replication origin interference increases the spacing between initiation events in human cells." Mol Biol Cell **17**(12): 5337-5345.
Lei, M., Y. Kawasaki, M. R. Young, M. Kihara, A. Sugino and B. K. Tye (1997). "Mcm2 is a target of regulation by Cdc7-Dbf4 during the initiation of DNA synthesis." Genes Dev **11**(24): 3365-3374.
Lejeune, E., M. Bortfeld, S. A. White, A. L. Pidoux, K. Ekwall, R. C. Allshire and A. G. Ladurner (2007). "The chromatin-remodeling factor FACT contributes to centromeric heterochromatin independently of RNAi." Curr Biol **17**(14): 1219-1224.
Li, F., D. B. Goto, M. Zaratiegui, X. Tang, R. Martienssen and W. Z. Cande (2005). "Two novel proteins, dos1 and dos2, interact with rik1 to regulate heterochromatic RNA interference and histone modification." Curr Biol **15**(16): 1448-1457.
Li, P. C., L. Chretien, J. Cote, T. J. Kelly and S. L. Forsburg (2011). "S. pombe replication protein Cdc18 (Cdc6) interacts with Swi6 (HP1) heterochromatin protein: region specific effects and replication timing in the centromere." Cell Cycle **10**(2): 323-336.
Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander and J. Dekker (2009). "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." Science **326**(5950): 289-293.
Lima-de-Faria, A. and H. Jaworska (1968). "Late DNA synthesis in heterochromatin." Nature **217**(5124): 138-142.
Liu, G., X. Chen, Y. Gao, T. Lewis, J. Barthelemy and M. Leffak (2012). "Altered replication in human cells promotes DMPK (CTG)(n) . (CAG)(n) repeat instability." Mol Cell Biol **32**(9): 1618-1632.
Lu, H., F. Giordano and Z. Ning (2016). "Oxford Nanopore MinION Sequencing and Genome Assembly." Genomics Proteomics Bioinformatics **14**(5): 265-279.
Mans, B. J., V. Anantharaman, L. Aravind and E. V. Koonin (2004). "Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex." Cell Cycle **3**(12): 1612-1637.
Marguerat, S., A. Schmidt, S. Codlin, W. Chen, R. Aebersold and J. Bahler (2012). "Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells." Cell **151**(3): 671-683.
Martienssen, R. and D. Moazed (2015). "RNAi and heterochromatin assembly." Cold Spring Harb Perspect Biol **7**(8): a019323.
Masai, H., E. Matsui, Z. You, Y. Ishimi, K. Tamai and K. Arai (2000). "Human Cdc7-related kinase complex. In vitro phosphorylation of MCM by concerted actions of Cdks and Cdc7 and that of a criticial threonine residue of Cdc7 bY Cdks." J Biol Chem **275**(37): 29042-29052.
Masai, H., S. Matsumoto, Z. You, N. Yoshizawa-Sugata and M. Oda (2010). "Eukaryotic chromosome DNA replication: where, when, and how?" Annu Rev Biochem **79**: 89-130.
Masai, H., C. C. Yang and S. Matsumoto (2017). "Mrc1/Claspin: a new role for regulation of origin firing." Curr Genet **63**(5): 813-818.
Matsumoto, S., Y. Kanoh, M. Shimmoto, M. Hayano, K. Ueda, R. Fukatsu, N. Kakusho and H. Masai (2017). "Checkpoint-Independent Regulation of Origin Firing by Mrc1 through Interaction with Hsk1 Kinase." Mol Cell Biol **37**(7).
Merrick, K. A. and R. P. Fisher (2012). "Why minimal is not optimal: driving the mammalian cell cycle--and drug discovery--with a physiologic CDK control network." Cell Cycle **11**(14): 2600-2605.

Miller, K. M., M. G. Ferreira and J. P. Cooper (2005). "Taz1, Rap1 and Rif1 act both interdependently and independently to maintain telomeres." EMBO J **24**(17): 3128-3135.
Miyabe, I., T. A. Kunkel and A. M. Carr (2011). "The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved." PLoS Genet **7**(12): e1002407.
Miyabe, I., K. Mizuno, A. Keszthelyi, Y. Daigaku, M. Skouteri, S. Mohebi, T. A. Kunkel, J. M. Murray and A. M. Carr (2015). "Polymerase delta replicates both strands after homologous recombination-dependent fork restart." Nat Struct Mol Biol **22**(11): 932-938.
Mizuguchi, T., G. Fudenberg, S. Mehta, J. M. Belton, N. Taneja, H. D. Folco, P. FitzGerald, J. Dekker, L. Mirny, J. Barrowman and S. I. S. Grewal (2014). "Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe." Nature **516**(7531): 432-435.
Mizuno, K., S. Lambert, G. Baldacci, J. M. Murray and A. M. Carr (2009). "Nearby inverted repeats fuse to generate acentric and dicentric palindromic chromosomes by a replication template exchange mechanism." Genes Dev **23**(24): 2876-2886.
Mohanty, B. K., N. K. Bairwa and D. Bastia (2006). "The Tof1p-Csm3p protein complex counteracts the Rrm3p helicase to control replication termination of Saccharomyces cerevisiae." Proc Natl Acad Sci U S A **103**(4): 897-902.
Moyer, S. E., P. W. Lewis and M. R. Botchan (2006). "Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase." Proc Natl Acad Sci U S A **103**(27): 10236-10241.
Muller, C. A., M. Hawkins, R. Retkute, S. Malla, R. Wilson, M. J. Blythe, R. Nakato, M. Komata, K. Shirahige, A. P. de Moura and C. A. Nieduszynski (2014). "The dynamics of genome replication using deep sequencing." Nucleic Acids Res **42**(1): e3.
Muramatsu, S., K. Hirai, Y. S. Tak, Y. Kamimura and H. Araki (2010). "CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol (epsilon}, and GINS in budding yeast." Genes Dev **24**(6): 602-612.
Nakagawa, H., J. K. Lee, J. Hurwitz, R. C. Allshire, J. Nakayama, S. I. Grewal, K. Tanaka and Y. Murakami (2002). "Fission yeast CENP-B homologs nucleate centromeric heterochromatin by promoting heterochromatin-specific histone tail modifications." Genes Dev **16**(14): 1766-1778.
Nakajima, R. and H. Masukata (2002). "SpSld3 is required for loading and maintenance of SpCdc45 on chromatin in DNA replication in fission yeast." Mol Biol Cell **13**(5): 1462-1472.
Nakaseko, Y., Y. Adachi, S. Funahashi, O. Niwa and M. Yanagida (1986). "Chromosome walking shows a highly homologous repetitive sequence present in all the centromere regions of fission yeast." EMBO J **5**(5): 1011-1021.
Nakaseko, Y., N. Kinoshita and M. Yanagida (1987). "A novel sequence common to the centromere regions of Schizosaccharomyces pombe chromosomes." Nucleic Acids Res **15**(12): 4705-4715.
Nakayama, J., A. J. Klar and S. I. Grewal (2000). "A chromodomain protein, Swi6, performs imprinting functions in fission yeast during mitosis and meiosis." Cell **101**(3): 307-317.
Nakayama, J., J. C. Rice, B. D. Strahl, C. D. Allis and S. I. Grewal (2001). "Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly." Science **292**(5514): 110-113.
Naumova, N., M. Imakaev, G. Fudenberg, Y. Zhan, B. R. Lajoie, L. A. Mirny and J. Dekker (2013). "Organization of the mitotic chromosome." Science **342**(6161): 948-953.
Negrini, S., V. G. Gorgoulis and T. D. Halazonetis (2010). "Genomic instability--an evolving hallmark of cancer." Nat Rev Mol Cell Biol **11**(3): 220-228.
Nick McElhinny, S. A., D. A. Gordenin, C. M. Stith, P. M. Burgers and T. A. Kunkel (2008). "Division of labor at the eukaryotic replication fork." Mol Cell **30**(2): 137-144.
Nishitani, H., N. Sugimoto, V. Roukos, Y. Nakanishi, M. Saijo, C. Obuse, T. Tsurimoto, K. I. Nakayama, K. Nakayama, M. Fujita, Z. Lygerou and T. Nishimoto (2006). "Two E3 ubiquitin ligases, SCF-Skp2 and DDB1-Cul4, target human Cdt1 for proteolysis." EMBO J **25**(5): 1126-1136.

Noguchi, E., C. Noguchi, L. L. Du and P. Russell (2003). "Swi1 prevents replication fork collapse and controls checkpoint kinase Cds1." Mol Cell Biol **23**(21): 7861-7874.
Noguchi, E., C. Noguchi, W. H. McDonald, J. R. Yates, 3rd and P. Russell (2004). "Swi1 and Swi3 are components of a replication fork protection complex in fission yeast." Mol Cell Biol **24**(19): 8342-8355.
Noguchi, E., P. Shanahan, C. Noguchi and P. Russell (2002). "CDK phosphorylation of Drc1 regulates DNA replication in fission yeast." Curr Biol **12**(7): 599-605.
Nora, E. P., B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Bluthgen, J. Dekker and E. Heard (2012). "Spatial partitioning of the regulatory landscape of the X-inactivation centre." Nature **485**(7398): 381-385.
O'Donnell, M., L. Langston and B. Stillman (2013). "Principles and concepts of DNA replication in bacteria, archaea, and eukarya." Cold Spring Harb Perspect Biol **5**(7).
O'Donnell, M. and H. Li (2016). "The Eukaryotic Replisome Goes Under the Microscope." Curr Biol **26**(6): R247-256.
Okonechnikov, K., A. Conesa and F. Garcia-Alcalde (2016). "Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data." Bioinformatics **32**(2): 292-294.
Ostrow, A. Z., R. Kalhor, Y. Gan, S. K. Villwock, C. Linke, M. Barberis, L. Chen and O. M. Aparicio (2017). "Conserved forkhead dimerization motif controls DNA replication timing and spatial organization of chromosomes in S. cerevisiae." Proc Natl Acad Sci U S A **114**(12): E2411-E2419.
Paek, A. L., S. Kaochar, H. Jones, A. Elezaby, L. Shanks and T. Weinert (2009). "Fusion of nearby inverted repeats by a replication-based mechanism leads to formation of dicentric and acentric chromosomes that cause genome instability in budding yeast." Genes Dev **23**(24): 2861-2875.
Park, S., E. E. Patterson, J. Cobb, A. Audhya, M. R. Gartenberg and C. A. Fox (2011). "Palmitoylation controls the dynamics of budding-yeast heterochromatin via the telomere-binding protein Rif1." Proc Natl Acad Sci U S A **108**(35): 14572-14577.
Partridge, J. F., B. Borgstrom and R. C. Allshire (2000). "Distinct protein interaction domains and protein spreading in a complex centromere." Genes Dev **14**(7): 783-791.
Partridge, J. F., K. S. Scott, A. J. Bannister, T. Kouzarides and R. C. Allshire (2002). "cis-acting DNA from fission yeast centromeres mediates histone H3 methylation and recruitment of silencing factors and cohesin to an ectopic site." Curr Biol **12**(19): 1652-1660.
Passarge, E. (1979). "Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago." Am J Hum Genet **31**(2): 106-115.
Patel, P. K., B. Arcangioli, S. P. Baker, A. Bensimon and N. Rhind (2006). "DNA replication origins fire stochastically in fission yeast." Mol Biol Cell **17**(1): 308-316.
Peace, J. M., A. Ter-Zakarian and O. M. Aparicio (2014). "Rif1 regulates initiation timing of late replication origins throughout the S. cerevisiae genome." PLoS One **9**(5): e98501.
Perez-Arnaiz, P., I. Bruck and D. L. Kaplan (2016). "Mcm10 coordinates the timely assembly and activation of the replication fork helicase." Nucleic Acids Res **44**(1): 315-329.
Peric-Hupkes, D., W. Meuleman, L. Pagie, S. W. Bruggeman, I. Solovei, W. Brugman, S. Graf, P. Flicek, R. M. Kerkhoven, M. van Lohuizen, M. Reinders, L. Wessels and B. van Steensel (2010). "Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation." Mol Cell **38**(4): 603-613.
Pichugina, T., T. Sugawara, A. Kaykov, W. Schierding, K. Masuda, J. Uewaki, R. S. Grand, J. R. Allison, R. A. Martienssen, P. Nurse, M. Ueno and J. M. O'Sullivan (2016). "A diffusion model for the coordination of DNA replication in Schizosaccharomyces pombe." Sci Rep **6**: 18757.
Pickersgill, H., B. Kalverda, E. de Wit, W. Talhout, M. Fornerod and B. van Steensel (2006). "Characterization of the Drosophila melanogaster genome at the nuclear lamina." Nat Genet **38**(9): 1005-1014.
Pope, B. D., T. Chandra, Q. Buckley, M. Hoare, T. Ryba, F. K. Wiseman, A. Kuta, M. D. Wilson, D. T. Odom and D. M. Gilbert (2012). "Replication-timing boundaries facilitate cell-type and

species-specific regulation of a rearranged human chromosome in mouse." Hum Mol Genet **21**(19): 4162-4170.
Pope, B. D., T. Ryba, V. Dileep, F. Yue, W. Wu, O. Denas, D. L. Vera, Y. Wang, R. S. Hansen, T. K. Canfield, R. E. Thurman, Y. Cheng, G. Gulsoy, J. H. Dennis, M. P. Snyder, J. A. Stamatoyannopoulos, J. Taylor, R. C. Hardison, T. Kahveci, B. Ren and D. M. Gilbert (2014). "Topologically associating domains are stable units of replication-timing regulation." Nature **515**(7527): 402-405.
Raghuraman, M. K., E. A. Winzeler, D. Collingwood, S. Hunt, L. Wodicka, A. Conway, D. J. Lockhart, R. W. Davis, B. J. Brewer and W. L. Fangman (2001). "Replication dynamics of the yeast genome." Science **294**(5540): 115-121.
Ralph, E., E. Boye and S. E. Kearsey (2006). "DNA damage induces Cdt1 proteolysis in fission yeast through a pathway dependent on Cdt2 and Ddb1." EMBO Rep **7**(11): 1134-1139.
Randell, J. C., J. L. Bowers, H. K. Rodriguez and S. P. Bell (2006). "Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase." Mol Cell **21**(1): 29-39.
Regnier, V., P. Vagnarelli, T. Fukagawa, T. Zerjal, E. Burns, D. Trouche, W. Earnshaw and W. Brown (2005). "CENP-A is required for accurate chromosome segregation and sustained kinetochore association of BubR1." Mol Cell Biol **25**(10): 3967-3981.
Reinhart, B. J. and D. P. Bartel (2002). "Small RNAs correspond to centromere heterochromatic repeats." Science **297**(5588): 1831.
Retkute, R., C. A. Nieduszynski and A. de Moura (2012). "Mathematical modeling of genome replication." Phys Rev E Stat Nonlin Soft Matter Phys **86**(3 Pt 1): 031916.
Rhind, N. (2006). "DNA replication timing: random thoughts about origin firing." Nat Cell Biol **8**(12): 1313-1316.
Rhind, N. and D. M. Gilbert (2013). "DNA replication timing." Cold Spring Harb Perspect Biol **5**(8): a010132.
Rhind, N., S. C. Yang and J. Bechhoefer (2010). "Reconciling stochastic origin firing with defined replication timing." Chromosome Res **18**(1): 35-43.
Rivera-Mulia, J. C. and D. M. Gilbert (2016). "Replication timing and transcriptional control: beyond cause and effect-part III." Curr Opin Cell Biol **40**: 168-178.
Rowlands, H., P. Dhavarasa, A. Cheng and K. Yankulov (2017). "Forks on the Run: Can the Stalling of DNA Replication Promote Epigenetic Changes?" Front Genet **8**: 86.
Rozenzhak, S., E. Mejia-Ramirez, J. S. Williams, L. Schaffer, J. A. Hammond, S. R. Head and P. Russell (2010). "Rad3 decorates critical chromosomal domains with gammaH2A to protect genome integrity during S-Phase in fission yeast." PLoS Genet **6**(7): e1001032.
Ryba, T., D. Battaglia, B. D. Pope, I. Hiratani and D. M. Gilbert (2011). "Genome-scale analysis of replication timing: from bench to bioinformatics." Nat Protoc **6**(6): 870-895.
Ryba, T., I. Hiratani, J. Lu, M. Itoh, M. Kulik, J. Zhang, T. C. Schulz, A. J. Robins, S. Dalton and D. M. Gilbert (2010). "Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types." Genome Res **20**(6): 761-770.
Sabouri, N., J. A. Capra and V. A. Zakian (2014). "The essential Schizosaccharomyces pombe Pfh1 DNA helicase promotes fork movement past G-quadruplex motifs to prevent DNA damage." BMC Biol **12**: 101.
Sadaie, M., T. Iida, T. Urano and J. Nakayama (2004). "A chromodomain protein, Chp1, is required for the establishment of heterochromatin in fission yeast." EMBO J **23**(19): 3825-3835.
Saksouk, N., E. Simboeck and J. Dejardin (2015). "Constitutive heterochromatin formation and transcription in mammals." Epigenetics Chromatin **8**: 3.
Sawyer, S. L., I. H. Cheng, W. Chai and B. K. Tye (2004). "Mcm10 and Cdc45 cooperate in origin activation in Saccharomyces cerevisiae." J Mol Biol **340**(2): 195-202.
Schalbetter, S. A., A. Goloborodko, G. Fudenberg, J. M. Belton, C. Miles, M. Yu, J. Dekker, L. Mirny and J. Baxter (2017). "SMC complexes differentially compact mitotic chromosomes according to genomic context." Nat Cell Biol **19**(9): 1071-1080.

Sexton, T. and G. Cavalli (2015). "The role of chromosome domains in shaping the functional genome." Cell **160**(6): 1049-1059.
Sheu, Y. J. and B. Stillman (2010). "The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4." Nature **463**(7277): 113-117.
Shimmoto, M., S. Matsumoto, Y. Odagiri, E. Noguchi, P. Russell and H. Masai (2009). "Interactions between Swi1-Swi3, Mrc1 and S phase kinase, Hsk1 may regulate cellular responses to stalled replication forks in fission yeast." Genes Cells **14**(6): 669-682.
Shishkin, A. A., I. Voineagu, R. Matera, N. Cherng, B. T. Chernet, M. M. Krasilnikova, V. Narayanan, K. S. Lobachev and S. M. Mirkin (2009). "Large-scale expansions of Friedreich's ataxia GAA repeats in yeast." Mol Cell **35**(1): 82-92.
Sicard, H., M. Faubladier, J. Noaillac-Depeyre, I. Leger-Silvestre, N. Gas and M. Caizergues-Ferrer (1998). "The role of the Schizosaccharomyces pombe gar2 protein in nucleolar structure and function depends on the concerted action of its highly charged N terminus and its RNA-binding domains." Mol Biol Cell **9**(8): 2011-2023.
Sparks, J. L., H. Chon, S. M. Cerritelli, T. A. Kunkel, E. Johansson, R. J. Crouch and P. M. Burgers (2012). "RNase H2-initiated ribonucleotide excision repair." Mol Cell **47**(6): 980-986.
Steglich, B., G. J. Filion, B. van Steensel and K. Ekwall (2012). "The inner nuclear membrane proteins Man1 and Ima1 link to two different types of chromatin at the nuclear periphery in S. pombe." Nucleus **3**(1): 77-87.
Sugiyama, T., H. P. Cam, R. Sugiyama, K. Noma, M. Zofall, R. Kobayashi and S. I. Grewal (2007). "SHREC, an effector complex for heterochromatic transcriptional silencing." Cell **128**(3): 491-504.
Sugiyama, T., G. Thillainadesan, V. R. Chalamcharla, Z. Meng, V. Balachandran, J. Dhakshnamoorthy, M. Zhou and S. I. S. Grewal (2016). "Enhancer of Rudimentary Cooperates with Conserved RNA-Processing Factors to Promote Meiotic mRNA Decay and Facultative Heterochromatin Assembly." Mol Cell **61**(5): 747-759.
Takahashi, K., S. Murakami, Y. Chikashige, H. Funabiki, O. Niwa and M. Yanagida (1992). "A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere." Mol Biol Cell **3**(7): 819-835.
Tanae, K., T. Horiuchi, Y. Matsuo, S. Katayama and M. Kawamukai (2012). "Histone chaperone Asf1 plays an essential role in maintaining genomic stability in fission yeast." PLoS One **7**(1): e30472.
Tanaka, K. and P. Russell (2001). "Mrc1 channels the DNA replication arrest signal to checkpoint kinase Cds1." Nat Cell Biol **3**(11): 966-972.
Tashiro, S., T. Asano, J. Kanoh and F. Ishikawa (2013). "Transcription-induced chromatin association of RNA surveillance factors mediates facultative heterochromatin formation in fission yeast." Genes Cells **18**(4): 327-339.
Tazumi, A., M. Fukuura, R. Nakato, A. Kishimoto, T. Takenaka, S. Ogawa, J. H. Song, T. S. Takahashi, T. Nakagawa, K. Shirahige and H. Masukata (2012). "Telomere-binding protein Taz1 controls global replication timing through its localization near late replication origins in fission yeast." Genes Dev **26**(18): 2050-2062.
Thon, G., K. R. Hansen, S. P. Altes, D. Sidhu, G. Singh, J. Verhein-Hansen, M. J. Bonaduce and A. J. Klar (2005). "The Clr7 and Clr8 directionality factors and the Pcu4 cullin mediate heterochromatin formation in the fission yeast Schizosaccharomyces pombe." Genetics **171**(4): 1583-1595.
Toteva, T., B. Mason, Y. Kanoh, P. Brogger, D. Green, J. Verhein-Hansen, H. Masai and G. Thon (2017). "Establishment of expression-state boundaries by Rif1 and Taz1 in fission yeast." Proc Natl Acad Sci U S A **114**(5): 1093-1098.
Tourriere, H., G. Versini, V. Cordon-Preciado, C. Alabert and P. Pasero (2005). "Mrc1 and Tof1 promote replication fork progression and recovery independently of Rad53." Mol Cell **19**(5): 699-706.

Tsukada, Y., J. Fang, H. Erdjument-Bromage, M. E. Warren, C. H. Borchers, P. Tempst and Y. Zhang (2006). "Histone demethylation by a family of JmjC domain-containing proteins." Nature **439**(7078): 811-816.
Uzawa, S. and M. Yanagida (1992). "Visualization of centromeric and nucleolar DNA in fission yeast by fluorescence in situ hybridization." J Cell Sci **101 (Pt 2)**: 267-275.
Van Hooser, A. A., Ouspenski, II, H. C. Gregson, D. A. Starr, T. J. Yen, M. L. Goldberg, K. Yokomori, W. C. Earnshaw, K. F. Sullivan and B. R. Brinkley (2001). "Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A." J Cell Sci **114**(Pt 19): 3529-3542.
van Steensel, B., J. Delrow and S. Henikoff (2001). "Chromatin profiling using targeted DNA adenine methyltransferase." Nat Genet **27**(3): 304-308.
Verdel, A., S. Jia, S. Gerber, T. Sugiyama, S. Gygi, S. I. Grewal and D. Moazed (2004). "RNAi-mediated targeting of heterochromatin by the RITS complex." Science **303**(5658): 672-676.
Voineagu, I., V. Narayanan, K. S. Lobachev and S. M. Mirkin (2008). "Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins." Proc Natl Acad Sci U S A **105**(29): 9936-9941.
Voineagu, I., C. F. Surka, A. A. Shishkin, M. M. Krasilnikova and S. M. Mirkin (2009). "Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility." Nat Struct Mol Biol **16**(2): 226-228.
Volpe, T. and R. A. Martienssen (2011). "RNA interference and heterochromatin assembly." Cold Spring Harb Perspect Biol **3**(9): a003731.
Volpe, T. A., C. Kidner, I. M. Hall, G. Teng, S. I. Grewal and R. A. Martienssen (2002). "Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi." Science **297**(5588): 1833-1837.
Wang, J., S. T. Jia and S. Jia (2016). "New Insights into the Regulation of Heterochromatin." Trends Genet **32**(5): 284-294.
Wang, J., X. Tadeo, H. Hou, P. G. Tu, J. Thompson, J. R. Yates, 3rd and S. Jia (2013). "Epe1 recruits BET family bromodomain protein Bdf2 to establish heterochromatin boundaries." Genes Dev **27**(17): 1886-1902.
Watson, A. T., V. Garcia, N. Bone, A. M. Carr and J. Armstrong (2008). "Gene tagging and gene replacement using recombinase-mediated cassette exchange in Schizosaccharomyces pombe." Gene **407**(1-2): 63-74.
Winkler, D. D., U. M. Muthurajan, A. R. Hieb and K. Luger (2011). "Histone chaperone FACT coordinates nucleosome interaction through multiple synergistic binding events." J Biol Chem **286**(48): 41883-41892.
Wohlschlegel, J. A., S. K. Dhar, T. A. Prokhorova, A. Dutta and J. C. Walter (2002). "Xenopus Mcm10 binds to origins of DNA replication after Mcm2-7 and stimulates origin binding of Cdc45." Mol Cell **9**(2): 233-240.
Wohlschlegel, J. A., B. T. Dwyer, S. K. Dhar, C. Cvetic, J. C. Walter and A. Dutta (2000). "Inhibition of eukaryotic DNA replication by geminin binding to Cdt1." Science **290**(5500): 2309-2312.
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. Schafer, S. Muller-Auer, C. Gabel, M. Fuchs, A. Dusterhoft, 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. Dreano, 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. Sanchez, F. del Rey, J. Benito, A. Dominguez, J. L. Revuelta, S. Moreno, J. Armstrong, S. L. Forsburg, L. Cerutti, 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): 871-880.
Wotton, D. and D. Shore (1997). "A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in Saccharomyces cerevisiae." Genes Dev **11**(6): 748-760.
Xu, J., Y. Yanagisawa, A. M. Tsankov, C. Hart, K. Aoki, N. Kommajosyula, K. E. Steinmann, J. Bochicchio, C. Russ, A. Regev, O. J. Rando, C. Nusbaum, H. Niki, P. Milos, Z. Weng and N. Rhind (2012). "Genome-wide identification and characterization of replication origins by deep sequencing." Genome Biol **13**(4): R27.
Yabuuchi, H., Y. Yamada, T. Uchida, T. Sunathvanichkul, T. Nakagawa and H. Masukata (2006). "Ordered assembly of Sld3, GINS and Cdc45 is distinctly regulated by DDK and CDK for activation of replication origins." EMBO J **25**(19): 4663-4674.
Yaffe, E., S. Farkash-Amar, A. Polten, Z. Yakhini, A. Tanay and I. Simon (2010). "Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture." PLoS Genet **6**(7): e1001011.
Yang, S. C., N. Rhind and J. Bechhoefer (2010). "Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing." Mol Syst Biol **6**: 404.
Yeeles, J. T., T. D. Deegan, A. Janska, A. Early and J. F. Diffley (2015). "Regulated eukaryotic DNA replication origin firing with purified proteins." Nature **519**(7544): 431-435.
Yeeles, J. T. P., A. Janska, A. Early and J. F. X. Diffley (2017). "How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication." Mol Cell **65**(1): 105-116.
Yurieva, O. and M. O'Donnell (2016). "Reconstitution of a eukaryotic replisome reveals the mechanism of asymmetric distribution of DNA polymerases." Nucleus **7**(4): 360-368.
Zaaijer, S., N. Shaikh, R. K. Nageshan and J. P. Cooper (2016). "Rif1 Regulates the Fate of DNA Entanglements during Mitosis." Cell Rep **16**(1): 148-160.
Zegerman, P. and J. F. Diffley (2007). "Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast." Nature **445**(7125): 281-285.
Zegerman, P. and J. F. Diffley (2009). "DNA replication as a target of the DNA damage checkpoint." DNA Repair (Amst) **8**(9): 1077-1088.
Zhai, Y., N. Li, H. Jiang, X. Huang, N. Gao and B. K. Tye (2017). "Unique Roles of the Non-identical MCM Subunits in DNA Replication Licensing." Mol Cell **67**(2): 168-179.
Zhang, J. and J. C. Walter (2014). "Mechanism and regulation of incisions during DNA interstrand cross-link repair." DNA Repair (Amst) **19**: 135-142.
Zhong, W., H. Feng, F. E. Santiago and E. T. Kipreos (2003). "CUL-4 ubiquitin ligase maintains genome stability by restraining DNA-replication licensing." Nature **423**(6942): 885-889.
Zimmermann, M. and T. de Lange (2014). "53BP1: pro choice in DNA repair." Trends Cell Biol **24**(2): 108-117.
Zink, D. (2006). "The temporal program of DNA replication: new insights into old questions." Chromosoma **115**(4): 273-287.
Zofall, M. and S. I. Grewal (2006). "Swi6/HP1 recruits a JmjC domain protein to facilitate transcription of heterochromatic repeats." Mol Cell **22**(5): 681-692.
Zofall, M., D. R. Smith, T. Mizuguchi, J. Dhakshnamoorthy and S. I. S. Grewal (2016). "Taz1-Shelterin Promotes Facultative Heterochromatin Assembly at Chromosome-Internal Sites Containing Late Replication Origins." Mol Cell **62**(6): 862-874.
Zofall, M., S. Yamanaka, F. E. Reyes-Turcu, K. Zhang, C. Rubin and S. I. Grewal (2012). "RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation." Science **335**(6064): 96-100.

Appendix

9.1 Pu-Seq R script

Script was written by Dr. Andrea Keszthelyi. Thresholds used are indicated in bold

```
sample<-"NAME"

dat_df<- read.csv("path/rNTP-pol-d-",sample,".e1.f-w300.count.csv", sep=""))
dat_dr<- read.csv("path/rNTP-pol-d-",sample,".e1.r-w300.count.csv", sep=""))
dat_ef<- read.csv("path/rNTP-pol-e-",sample,".e1.f-w300.count.csv", sep=""))
dat_er<- read.csv("path/rNTP-pol-e-",sample,".e1.r-w300.count.csv", sep=""))

N<-3
NN<-3
p<-0.3
name<-sample
chromoname_in<-"Chromosome_"
outpath<-""
################################################  Functions
#################################################################

moving.ave.v2 <- function(data, n){ # subroutine to calculate moving averages

  dataN <- length(data)

  start <- c(rep(1,n), 1:(dataN-n))
  end   <- c((n+1):dataN, rep(dataN,n))
  se <- cbind(start, end)

  average.se <- function(n)mean(data[n[1]:n[2]][!is.nan(data[n[1]:n[2]])])
  r <- apply(se, 1, average.se)

  (r)
}

diff.sequence <- function(vec){
  diff <-c()
  diff[1]=0;
  for(i in 2:length(vec)){
    diff[i] = vec[i]-vec[i-1]
  }
  return(diff)
}

Findlocalmax<-function(diffdata,position,percentile){

  max<-c(which(diff(c(TRUE,diff(diffdata)>=0,FALSE))<0 & diffdata>0) )
  tableall<-cbind(position[max], diffdata[max])
  perc<-quantile(tableall[,2],percentile)
  per<-which(tableall[,2]<=perc)
  table<-tableall[-per,]
  return(table)
}

Closeori<-function(pos, bin){

  x<-c(1:length(pos))
  remove<-c()
  replace<-c()
  for(i in 1:(length(pos)-1)){
    if (abs(pos[i]-pos[i+1])==bin)
    {remove<-c(remove,x[i])
    next
    }else if
    (abs(pos[i]-pos[i+1])==2*bin)
    {remove<-c(remove,x[i],x[i+1])
    replace<-c(replace,(pos[x[i]]+bin))
    next
    }else if
    (abs(pos[i]-pos[i+1])==3*bin)
    {remove<-c(remove,x[i],x[i+1])
    replace<-c(replace,(pos[x[i]]+bin))
    next
    }else if
    (abs(pos[i]-pos[i+1])==4*bin)
```

```
    {remove<-c(remove,x[i],x[i+1])
    replace<-c(replace,(pos[x[i]]+2*bin))
    }

  }
  if(length(remove>0)){
    posremove<-c(pos[-remove])}else
      posremove<-pos
  posreplace<-sort(unique(c(posremove,replace)))
  return(posreplace)


}


#tableclose=peak positions from diff, ratio=pol usage ratio, pos=all position)
orieff<-function(close, ratio, pos){

  maxpos<-c(match(close,pos))

  ratiomin<-c(tail((which(diff(c(FALSE,diff(ratio[1:maxpos[1]])>0,TRUE))>0)),n=1))

  for (i in 1:(length(maxpos)-1)){

    ratiomin<-c(ratiomin,(-
1+maxpos[i]+tail((which(diff(c(FALSE,diff(ratio[maxpos[i]:maxpos[i+1]])>0,TRUE))>0)),n=1
)))
  }

  ratiomax<-c()

  for (i in 1:(length(maxpos)-1)){
    ratiomax<-c(ratiomax,(-
1+maxpos[i]+head((which(diff(c(TRUE,diff(ratio[maxpos[i]:maxpos[i+1]])>=0,FALSE))<0)),n=
1)))
  }
  ratiomax<-c(ratiomax, (-
1+maxpos[length(maxpos)]+head((which(diff(c(TRUE,diff(ratio[maxpos[length(maxpos)]:lengt
h(ratio)])>=0,FALSE))<0)),n=1)))
  #which(diff(c(TRUE,diff(x)>=0,FALSE))<0)

  orieff<-c((ratio[ratiomax]-ratio[ratiomin])*100)
  oriefftable<-cbind(close,orieff)

  return(oriefftable)
}


orieff_merge<-function(orieff_ef, orieff_ef_pos, orieff_dr, orieff_dr_pos,chro,bin){

  value<-c()
  valuepos<-c()
  drpaired<-c()
  efpaired<-c()
  for (i in 1:length(orieff_ef_pos)){
    if
    (orieff_ef_pos[i] %in% orieff_dr_pos){
      value<-c(value, mean(c(orieff_ef[i],
orieff_dr[match(orieff_ef_pos[i],orieff_dr_pos)])))
      valuepos<-c(valuepos, orieff_ef_pos[i])
      drpaired<-c(drpaired, orieff_dr_pos[match(orieff_ef_pos[i],orieff_dr_pos)] )
      efpaired<-c(efpaired, orieff_ef_pos[i])
      next
    }else if
    ((orieff_ef_pos[i]+bin) %in% orieff_dr_pos){
      value<-c(value, mean(c(orieff_ef[i],
orieff_dr[(match((orieff_ef_pos[i]+bin),orieff_dr_pos))])))
      valuepos<-c(valuepos, orieff_ef_pos[i])
      drpaired<-c(drpaired, (orieff_dr_pos[match((orieff_ef_pos[i]+bin),orieff_dr_pos)]
))
      efpaired<-c(efpaired, orieff_ef_pos[i])
      next

    }else if
    ((orieff_ef_pos[i]-bin) %in% orieff_dr_pos){
      value<-c(value, mean(c(orieff_ef[i], orieff_dr[(match((orieff_ef_pos[i]-
bin),orieff_dr_pos))])))
```

```
      valuepos<-c(valuepos, orieff_ef_pos[i])
      drpaired<-c(drpaired, (orieff_dr_pos[match((orieff_ef_pos[i]-bin),orieff_dr_pos)]
))
      efpaired<-c(efpaired, orieff_ef_pos[i])
      next
    }else if
    ((orieff_ef_pos[i]+2*bin) %in% orieff_dr_pos){
      value<-c(value, mean(c(orieff_ef[i],
orieff_dr[(match((orieff_ef_pos[i]+2*bin),orieff_dr_pos))])))
      valuepos<-c(valuepos, (orieff_ef_pos[i]+bin))
      drpaired<-c(drpaired,
(orieff_dr_pos[match((orieff_ef_pos[i]+2*bin),orieff_dr_pos)] ))
      efpaired<-c(efpaired, orieff_ef_pos[i])
      next
    }else if
    ((orieff_ef_pos[i]-2*bin) %in% orieff_dr_pos){
      value<-c(value, mean(c(orieff_ef[i], orieff_dr[(match((orieff_ef_pos[i]-
2*bin),orieff_dr_pos))])))
      valuepos<-c(valuepos, (orieff_ef_pos[i]-bin))
      drpaired<-c(drpaired, (orieff_dr_pos[match((orieff_ef_pos[i]-
2*bin),orieff_dr_pos)] ))
      efpaired<-c(efpaired, orieff_ef_pos[i])
      next
    }else if
    ((orieff_ef_pos[i]+3*bin) %in% orieff_dr_pos){
      value<-c(value, mean(c(orieff_ef[i],
orieff_dr[(match((orieff_ef_pos[i]+3*bin),orieff_dr_pos))])))
      valuepos<-c(valuepos, (orieff_ef_pos[i]+2*bin))
      drpaired<-c(drpaired,
(orieff_dr_pos[match((orieff_ef_pos[i]+3*bin),orieff_dr_pos)] ))
      efpaired<-c(efpaired, orieff_ef_pos[i])
      next

    }else if
    ((orieff_ef_pos[i]-3*bin) %in% orieff_dr_pos){
      value<-c(value, mean(c(orieff_ef[i], orieff_dr[(match((orieff_ef_pos[i]-
3*bin),orieff_dr_pos))])))
      valuepos<-c(valuepos, (orieff_ef_pos[i]-2*bin))
      drpaired<-c(drpaired, (orieff_dr_pos[match((orieff_ef_pos[i]-
3*bin),orieff_dr_pos)] ))
      efpaired<-c(efpaired, orieff_ef_pos[i])
      next
    }else if
    ((orieff_ef_pos[i]+4*bin) %in% orieff_dr_pos){
      value<-c(value, mean(c(orieff_ef[i],
orieff_dr[(match((orieff_ef_pos[i]+4*bin),orieff_dr_pos))])))
      valuepos<-c(valuepos, (orieff_ef_pos[i]+2*bin))
      drpaired<-c(drpaired,
(orieff_dr_pos[match((orieff_ef_pos[i]+4*bin),orieff_dr_pos)] ))
      efpaired<-c(efpaired, orieff_ef_pos[i])
      next

    }else if
    ((orieff_ef_pos[i]-4*bin) %in% orieff_dr_pos){
      value<-c(value, mean(c(orieff_ef[i], orieff_dr[(match((orieff_ef_pos[i]-
4*bin),orieff_dr_pos))])))
      valuepos<-c(valuepos, (orieff_ef_pos[i]-2*bin))
      drpaired<-c(drpaired, (orieff_dr_pos[match((orieff_ef_pos[i]-
4*bin),orieff_dr_pos)] ))
      efpaired<-c(efpaired, orieff_ef_pos[i])

    }
  }

  efpairedno<-match(efpaired,orieff_ef_pos)
  drpairedno<-match(drpaired,orieff_dr_pos)
  efunpaired<-orieff_ef_pos[-efpairedno]
  drunpaired<-orieff_dr_pos[-drpairedno]

  if((length(efunpaired) != 0) & (length(drunpaired) != 0)){



    for (i in 1:length(drunpaired)){
      if
      (drunpaired[i] %in% efunpaired){
```

```
        value<-c(value, mean(c(orieff_dr[match(drunpaired[i],orieff_dr_pos)],
orieff_ef[match(drunpaired[i],orieff_ef_pos)])))
        valuepos<-c(valuepos, drunpaired[i])
        next
      }else if
      ((drunpaired[i]+bin) %in% efunpaired){
        value<-c(value, mean(c(orieff_dr[match(drunpaired[i],orieff_dr_pos)],
orieff_ef[match((drunpaired[i]+bin),orieff_ef_pos)])))
        valuepos<-c(valuepos, drunpaired[i])
        next

      }else if
      ((drunpaired[i]-bin) %in% efunpaired){
        value<-c(value, mean(c(orieff_dr[match(drunpaired[i],orieff_dr_pos)],
orieff_ef[match((drunpaired[i]-bin),orieff_ef_pos)])))
        valuepos<-c(valuepos, drunpaired[i])
        next

      }else if
      ((drunpaired[i]+2*bin) %in% efunpaired){
        value<-c(value, mean(c(orieff_dr[match(drunpaired[i],orieff_dr_pos)],
orieff_ef[match((drunpaired[i]+2*bin),orieff_ef_pos)])))
        valuepos<-c(valuepos, (drunpaired[i]+bin))
        next

      }else if
      ((drunpaired[i]-2*bin) %in% efunpaired){
        value<-c(value, mean(c(orieff_dr[match(drunpaired[i],orieff_dr_pos)],
orieff_ef[match((drunpaired[i]-2*bin),orieff_ef_pos)])))
        valuepos<-c(valuepos, (drunpaired[i]-bin))
        next

      }else if
      ((drunpaired[i]+3*bin) %in% efunpaired){
        value<-c(value, mean(c(orieff_dr[match(drunpaired[i],orieff_dr_pos)],
orieff_ef[match((drunpaired[i]+3*bin),orieff_ef_pos)])))
        valuepos<-c(valuepos, (drunpaired[i]+2*bin))
        next

      }else if
      ((drunpaired[i]-3*bin) %in% efunpaired){
        value<-c(value, mean(c(orieff_dr[match(drunpaired[i],orieff_dr_pos)],
orieff_ef[match((drunpaired[i]-3*bin),orieff_ef_pos)])))
        valuepos<-c(valuepos, (drunpaired[i]-2*bin))
        next
      }else if
      ((drunpaired[i]+4*bin) %in% efunpaired){
        value<-c(value, mean(c(orieff_dr[match(drunpaired[i],orieff_dr_pos)],
orieff_ef[match((drunpaired[i]+4*bin),orieff_ef_pos)])))
        valuepos<-c(valuepos, (drunpaired[i]+2*bin))
        next

      }else if
      ((drunpaired[i]-4*bin) %in% efunpaired){
        value<-c(value, mean(c(orieff_dr[match(drunpaired[i],orieff_dr_pos)],
orieff_ef[match((drunpaired[i]-4*bin),orieff_ef_pos)])))
        valuepos<-c(valuepos, (drunpaired[i]-2*bin))

      }}
    chromosome<-rep(chro,length(valuepos))
    orilist<-as.data.frame(cbind(chromosome,valuepos,value))
    colnames(orilist)<-c("chromosome","maxpos","efficiency")

    #### Remove duplicates and zero values ####

    orilist_dupl<-which(duplicated(orilist$maxpos))
    if (length(orilist_dupl)>0){
      a<-c()
      for (i in 1:length(orilist_dupl)){
        a<-
c(a,(as.numeric(as.vector(orilist$efficiency[orilist_dupl[i]]))+as.numeric(as.vector(ori
list$efficiency[orilist_dupl[i]-1])))/2)
        levels(orilist$efficiency)<-c(levels(orilist$efficiency),a)}
      orilist$efficiency[orilist_dupl-1]<-a
      orilist<-orilist[-orilist_dupl,]
      orilist<-orilist[orilist$efficiency !=0,]}
```

```
  }else if(length(efunpaired)==0 | length(drunpaired)==0){

    chromosome<-rep(chro,length(valuepos))
    orilist<-as.data.frame(cbind(chromosome,valuepos,value))
    colnames(orilist)<-c("chromosome","maxpos","efficiency")

    #### Remove duplicates and zero values ####

    orilist_dupl<-which(duplicated(orilist$maxpos))
    if (length(orilist_dupl)>0){
      a<-c()
      for (i in 1:length(orilist_dupl)){
        a<-
c(a,(as.numeric(as.vector(orilist$efficiency[orilist_dupl[i]]))+as.numeric(as.vector(ori
list$efficiency[orilist_dupl[i]-1])))/2)
        levels(orilist$efficiency)<-c(levels(orilist$efficiency),a)}
      orilist$efficiency[orilist_dupl-1]<-a
      orilist<-orilist[-orilist_dupl,]
      orilist<-orilist[orilist$efficiency !=0,]}}



  return(orilist)
}


Wigdata <- function(name, ratio.table, row,  bin, color, h.line,chro,chromoname_in){


  header0 = paste('track type=wiggle_0 name="', name,
                  '" description="generated by Puseq_deltadelta_50_50_general.R A.
Keszthelyi & Y. Daigaku, 2015', date(),
                  '" visibility=full autoScale=off color=', color,
                  ' yLineOnOff=on yLineMark=', h.line,
                  ' priority=10', sep="")
  wiglist<- list()
  wiglist<-c(wiglist, header0)

  for(chromo in chro[1:length(chro)]){


    header = paste('fixedStep chrom=',paste(chromoname_in,gsub("[^0-9]",
"",chromo),sep=""),
                   ' step=', bin, ' span=', bin, sep="")

    ratio.table.chr = ratio.table[ratio.table$chromosome==chromo,]
    data.chr<-as.numeric(as.vector(ratio.table.chr[,row]))
    data.chr[is.na(data.chr)]<- 0
    wiglist<-c(wiglist,header,list(data.chr))}

  return(wiglist)

}


#####################################  Calculations
#############################################################

bin<-dat_df$pos[2]-dat_df$pos[1]

### normalization to all counts

dat_df.norm <- dat_df[,3]/sum(dat_df[,3])
dat_dr.norm <- dat_dr[,3]/sum(dat_dr[,3])
dat_ef.norm <- dat_ef[,3]/sum(dat_ef[,3])
dat_er.norm <- dat_er[,3]/sum(dat_er[,3])

### normalization using delta and epsilon both

dat_df.ratio <- dat_df.norm/(dat_df.norm+dat_ef.norm)
dat_dr.ratio <- dat_dr.norm/(dat_dr.norm+dat_er.norm)
dat_ef.ratio <- dat_ef.norm/(dat_df.norm+dat_ef.norm)
dat_er.ratio <- dat_er.norm/(dat_dr.norm+dat_er.norm)


### adjust using the mean of delta & epsilon ration
```

```
dat_df.ratio.mean <- dat_df.ratio*0.5/mean(dat_df.ratio,na.rm=TRUE)
dat_dr.ratio.mean <- dat_dr.ratio*0.5/mean(dat_dr.ratio,na.rm=TRUE)
dat_ef.ratio.mean <- dat_ef.ratio*0.5/mean(dat_ef.ratio,na.rm=TRUE)
dat_er.ratio.mean <- dat_er.ratio*0.5/mean(dat_er.ratio,na.rm=TRUE)

### usage in both strands ####

delta.both   <- (dat_df.ratio.mean + dat_dr.ratio.mean)/2
epsilon.both <- (dat_ef.ratio.mean + dat_er.ratio.mean)/2


table.ratio_1<-
cbind(as.character(dat_df[,1]),dat_df[,2],dat_df.ratio.mean,dat_dr.ratio.mean,dat_ef.rat
io.mean,dat_er.ratio.mean, delta.both,epsilon.both)
colnames(table.ratio_1)<-(c("chromosome","position","delta on forward","delta on
reverse","epsilon on forward","epsilon on reverse","delta both strand","epsilon both
strand"))



table.ratio<-data.frame()
chro<-unique(dat_df$chro)
for(chromo in chro[1:length(chro)]){

  table.ratio.chr<-table.ratio_1[table.ratio_1[,1]==chromo,]



  # N = input$N.ratio.num   # the parameter for moving ave (2N+1)

  dat_df.chr.ratio.mean.ma <- moving.ave.v2(as.numeric(as.vector(table.ratio.chr[,3])),
N)
  dat_dr.chr.ratio.mean.ma <- moving.ave.v2(as.numeric(as.vector(table.ratio.chr[,4])),
N)
  dat_ef.chr.ratio.mean.ma <- moving.ave.v2(as.numeric(as.vector(table.ratio.chr[,5])),
N)
  dat_er.chr.ratio.mean.ma <- moving.ave.v2(as.numeric(as.vector(table.ratio.chr[,6])),
N)
  delta.both.ma <- moving.ave.v2(as.numeric(as.vector(table.ratio.chr[,7])), N)
  epsilon.both.ma <- moving.ave.v2(as.numeric(as.vector(table.ratio.chr[,8])), N)



  ####### repl timing
  leftfork_table<-rbind(dat_er.chr.ratio.mean.ma,dat_df.chr.ratio.mean.ma)
  leftfork<-colMeans(leftfork_table)


  Trep<-c(43)
  for (i in 1:length(leftfork)-1){
    if (is.na(leftfork[i+1])){
    Trep<-c(Trep,(Trep[i]))
    }else{
    Trep<-c(Trep,(bin*(1-(2*leftfork[i+1]))/1500)+Trep[i])
  }}

  Trep_norm<-Trep/mean(Trep)


  ####### differencial ###

  dat_ef.chr.diff<-
moving.ave.v2(diff.sequence(as.numeric(as.vector(dat_ef.chr.ratio.mean.ma))),NN)
  dat_dr.chr.diff<-
moving.ave.v2(diff.sequence(as.numeric(as.vector(dat_dr.chr.ratio.mean.ma))),NN)

  table.ratio.chro<-
as.data.frame(cbind(as.character(table.ratio.chr[,1]),table.ratio.chr[,2],dat_df.chr.rat
io.mean.ma,dat_dr.chr.ratio.mean.ma,dat_ef.chr.ratio.mean.ma,dat_er.chr.ratio.mean.ma,
delta.both.ma,epsilon.both.ma,Trep,leftfork,Trep_norm,dat_ef.chr.diff,dat_dr.chr.diff))
  table.ratio<-rbind(table.ratio,table.ratio.chro)
}
colnames(table.ratio)<-(c("chromosome","position","delta on forward","delta on
reverse","epsilon on forward","epsilon on reverse","delta both strand","epsilon both
strand","Trep","leftward fork","Trep_norm","diff_ef","diff_dr"))
```

```
orilist<-data.frame()

for(chromo in chro[1:length(chro)]){

  ratio.table.chr = table.ratio[table.ratio$chromosome==chromo,]

  dat_ef.chr.diff<-
moving.ave.v2(diff.sequence(as.numeric(as.vector(ratio.table.chr[,5]))),NN)
  dat_dr.chr.diff<-
moving.ave.v2(diff.sequence(as.numeric(as.vector(ratio.table.chr[,4]))),NN)

  #find local maxima - p= percentile treshold


  dat_ef.chr.table<-
Findlocalmax(dat_ef.chr.diff,(as.numeric(as.vector(ratio.table.chr[,2]))),p)
  dat_dr.chr.table<-
Findlocalmax(dat_dr.chr.diff,(as.numeric(as.vector(ratio.table.chr[,2]))),p)


  #merge peaks within 4 bins

  dat_ef.chr.table.merged<-Closeori(as.numeric(as.vector(dat_ef.chr.table[,1])),bin)
  dat_dr.chr.table.merged<-Closeori(as.numeric(as.vector(dat_dr.chr.table[,1])),bin)

  #calculate efficiency

  dat_ef.chr.orieff<-
orieff(dat_ef.chr.table.merged,as.numeric(as.vector(ratio.table.chr[,5])),as.numeric(as.
vector(ratio.table.chr[,2])))
  dat_dr.chr.orieff<-
orieff(dat_dr.chr.table.merged,as.numeric(as.vector(ratio.table.chr[,4])),as.numeric(as.
vector(ratio.table.chr[,2])))

  #find peaks that are in both within plusminus 4 bins

  orilist.chr<-
orieff_merge(dat_ef.chr.orieff[,2],dat_ef.chr.orieff[,1],dat_dr.chr.orieff[,2],dat_dr.ch
r.orieff[,1],chromo,bin)
  orilist.chr_upper<-quantile(as.numeric(as.vector(orilist.chr[,3])),0.999)
  orilist.chr[,3]<-as.numeric(as.vector(orilist.chr[,3]))*100/orilist.chr_upper


  orilist<-rbind(orilist,orilist.chr)

}


########################################## File output
##############################################

##### histograms
fs = 2

ppi <- 300
png(paste(sample,"_",p,"_","orihist_100.png",sep=""), width=10*ppi, height=10*ppi,
res=ppi)


par(mfrow=c(1,1), omd=c(0, 1, 0.2, 0.9), plt=c(0.15, 0.95, 0, 1),lwd=2)
hist(orilist[,3],100,main="",yaxt = "n", xaxt = "n", ylim=c(0,120), xlim=c(0,110) ,
ylab="", xlab="",  xaxs="i")
axis(1, at=c(0, 20, 40, 60, 80, 100),labels=c(0, 20, 40, 60, 80, 100), las=1,
cex.axis=fs, lwd=2)
axis(2, at=c(0, 20, 40, 60, 80,100,120),labels=c(0, 20, 40, 60, 80, 100,120), las=1,
cex.axis=fs, lwd=2)
mtext(paste(name,", treshold:",p*100,", no of oris: ",length(orilist$maxpos), sep=""),
side=3, cex=2, line = 0)
mtext("Efficiency", side=1, cex=2, line = 3)
#mtext(paste(name), side=3, cex=2, line = 0)
mtext("Frequency", side=2, cex=2, line = 4)
dev.off()

png(paste(sample,"_",p,"_","orihist_5.png",sep=""), width=10*ppi, height=10*ppi,
res=ppi)
par(mfrow=c(1,1), omd=c(0, 1, 0.2, 0.9), plt=c(0.15, 0.95, 0, 1),lwd=2)
```

```
hist(orilist[,3],5,main="",yaxt = "n", xaxt = "n", ylim=c(0,1200), xlim=c(0,120) ,
ylab="", xlab="",  xaxs="i")
axis(1, at=c(0, 20, 40, 60, 80, 100, 120),labels=c(0, 20, 40, 60, 80, 100, 120), las=1,
cex.axis=fs, lwd=2)
axis(2, at=c(0, 200, 400, 600, 800,1000, 1200),labels=c(0, 200, 400, 600, 800,1000,
1200), las=1, cex.axis=fs, lwd=2)
mtext(paste(name,", treshold:",p*100,", no of oris: ",length(orilist$maxpos), sep=""),
side=3, cex=2, line = 0)
mtext("Efficiency", side=1, cex=2, line = 3)
#mtext(paste(name), side=3, cex=2, line = 0)
mtext("Frequency", side=2, cex=2, line = 5)
dev.off()

##### bedgraph and csv

orieff_bedgraph<-data.frame()

for(chromo in chro[1:length(chro)]){

  table.ori.bedgraph.chr = orilist[orilist$chromosome==chromo,]

  chromoname=paste(chromoname_in,gsub("[^0-9]", "",chromo),sep="")


  pos1<-table.ori.bedgraph.chr$maxpos
  pos2<-as.numeric(as.vector(pos1))+bin
  value<-table.ori.bedgraph.chr$efficiency
  Chromosome<-rep(chromoname, length(table.ori.bedgraph.chr$maxpos))
  orieff_bedgraph.chr<-data.frame(chro=Chromosome, start=pos1, end=pos2, value=value)
  colnames(orieff_bedgraph.chr) <-c("chro", "start", "end", "value")
  orieff_bedgraph<-rbind(orieff_bedgraph,orieff_bedgraph.chr)
}

df_wig<-Wigdata(paste(name,"_delta_forward",sep=""),
table.ratio,3,bin,"blue",0.5,chro,chromoname_in)
dr_wig<-Wigdata(paste(name,"_delta_reverse",sep=""),
table.ratio,4,bin,"blue",0.5,chro,chromoname_in)
ef_wig<-Wigdata(paste(name,"_epsilon_forward",sep=""),
table.ratio,5,bin,"red",0.5,chro,chromoname_in)
er_wig<-Wigdata(paste(name,"_epsilon_reverse",sep=""),
table.ratio,6,bin,"red",0.5,chro,chromoname_in)
dd_wig<-Wigdata(paste(name,"_delta on both strand",sep=""),
table.ratio,7,bin,"blue",0.5,chro,chromoname_in)
Trep_wig<-Wigdata(paste(name,"_Trep",sep=""),
table.ratio,9,bin,"blue",0.5,chro,chromoname_in)
leftfork_wig<-Wigdata(paste(name,"_leftward_moving_fork",sep=""),
table.ratio,10,bin,"blue",0.5,chro,chromoname_in)
Trep_norm_wig<-Wigdata(paste(name,"_Trep_norm_ave",sep=""),
table.ratio,11,bin,"blue",0.5,chro,chromoname_in)
diff_ef_wig<-Wigdata(paste(name,"_diff_ef",sep=""), table.ratio,12, bin, "black",0.5,
chro, chromoname_in)
diff_dr_wig<-Wigdata(paste(name,"_diff_dr",sep=""), table.ratio,13, bin, "black",0.5,
chro, chromoname_in)


write.csv(table.ratio, file=paste(outpath,name,"_","all_ratios", '.csv', sep=''),
row.names = FALSE)
write.csv(orilist, file=paste(outpath,name,"_",p,"_","ori_table_perc", '.csv', sep=''),
row.names = FALSE)
write.table(orieff_bedgraph, file=paste(outpath,name,"_",p,"_","origins_perc",
'.bedgraph', sep=''), append = F, quote=F, col.names=F, row.names=F)


out_df<- file(paste(outpath,name,"_","delta_forward_wig", '.wig', sep=''), "w")
writeLines(unlist(df_wig),out_df,sep="\n")
close(out_df)


out_dr<- file(paste(outpath,name,"_","delta_reverse", '.wig', sep=''), "w")
writeLines(unlist(dr_wig),out_dr,sep="\n")
close(out_dr)

out_ef<- file(paste(outpath,name,"_","epsilon_forward", '.wig', sep=''), "w")
writeLines(unlist(ef_wig),out_ef,sep="\n")
close(out_ef)

out_er<- file(paste(outpath,name,"_","epsilon_reverse", '.wig', sep=''), "w")
```

```
writeLines(unlist(er_wig),out_er,sep="\n")
close(out_er)

out_dd<- file(paste(outpath,name,"_","delta_delta", '.wig', sep=''), "w")
writeLines(unlist(dd_wig),out_dd,sep="\n")
close(out_dd)


out_Trep<- file(paste(outpath,name,"_","Trep", '.wig', sep=''), "w")
writeLines(unlist(Trep_wig),out_Trep,sep="\n")
close(out_Trep)

out_leftfork<- file(paste(outpath,name,"_","leftward moving fork", '.wig', sep=''), "w")
writeLines(unlist(leftfork_wig),out_leftfork,sep="\n")
close(out_leftfork)

out_Trep_norm<- file(paste(outpath,name,"_","Trep_norm_ave", '.wig', sep=''), "w")
writeLines(unlist(Trep_norm_wig),out_Trep_norm,sep="\n")
close(out_Trep_norm)

out_diff_ef<- file(paste(outpath,name,"_","diff_ef", '.wig', sep=''), "w")
writeLines(unlist(diff_ef_wig),out_diff_ef,sep="\n")
close(out_diff_ef)

out_diff_dr<- file(paste(outpath,name,"_","diff_dr", '.wig', sep=''), "w")
writeLines(unlist(diff_dr_wig),out_diff_dr,sep="\n")
close(out_diff_dr)
```