EPIGENETIC REGULATION OF DNA REPLICATION IN *DROSOPHILA MELANOGASTER*

Robin Leigh Armstrong

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Genetics and Molecular Biology.

Chapel Hill
2020

Approved by:

Robert J. Duronio

Jeanette G. Cook

Matthew L. Hirsch

David M. MacAlpine

Daniel J. McKay

ABSTRACT

Robin Leigh Armstrong: Epigenetic regulation of DNA replication
(Under the direction of Robert J. Duronio)

How origins of DNA replication are specified and activated in the context of an intact metazoan genome remains poorly understood. In contrast to *Saccharomyces cerevisiae,* replication initiation in metazoan genomes is not directed by well-defined sequence motifs. Rather, local chromatin environments have emerged as potential regulators of replication, yielding early and late replicating regions of the genome. Transcriptionally active, accessible euchromatin typically replicates early during S phase, whereas transcriptionally repressive, inaccessible heterochromatin typically replicates late. Current models of replication posit a stochastic process in which a higher density of specified origins in euchromatin compared to heterochromatin increases the probability of replication initiation, resulting in the earlier replication of euchromatin relative to heterochromatin. Despite strong genome-wide correlations between replication and chromatin, a true causal relationship between the two has yet to be determined. We investigated how chromatin organization impacts replication in *Drosophila* using our genetic platform in which endogenous histone genes are replaced with transgenic histone genes encoding mutations that prevent modification of specific histone residues. To explore the relationship between euchromatin and replication, we implemented a whole-genome sequencing method to produce genome-wide replication timing profiles. We analyzed the X Chromosome, which in *Drosophila* is 2-fold more transcriptionally active,

replicates earlier, and is hyper-acetylated at H4K16 in XY males relative to XX females. H4K16R mutation prevents transcriptional hyper-activation and earlier replication of the male X chromosome, consistent with the notion that transcription promotes early replication. To determine whether perturbation of heterochromatin affects late replication, we generated replication profiles from H3K9R mutant tissue. Despite well-known correlations between late replication and heterochromatin, perturbation of heterochromatin structure through H3K9R mutation does not result in large-scale changes in replication timing suggesting critical regulation beyond chromatin structure. To identify other contributors to replication timing control, we explored the relative contributions of cell lineage, cell cycle, and the *trans*-acting factor Rif1. We identified that cell lineage, rather than changes in cell cycle status, drive replication timing programs. Furthermore, Rif1 regulates replication timing in a tissue-specific manner supporting the notion that additional mechanisms beyond chromatin structure are key regulators of replication of metazoan genomes.

To Lily Rose.
The greatest gift of graduate school.

ACKNOWLEDGEMENTS

I could never have successfully completed this journey without the support of many incredible individuals. I would first like to thank Bob, for without his mentorship and guidance, I would not be half of the scientist that I am today. Being one of Bob's graduate students is truly special—we get to experience the amazing combination of an unbelievably talented scientific mentor and an incredibly supportive advocate. I'm forever grateful that Bob both trusted me and encouraged me to be independent in almost everything that I did in his lab. The skills that I have gained from him will only continue to propel my career forward. Finally, I couldn't be more thankful to Bob for accepting and supporting who I am as a scientist and as a person. The Duronio lab became my second family over the past five years.

I would also like to thank the members of my dissertation committee for their support and scientific feedback during my graduate career. Matt, you taught me to thoroughly think through every experiment before I performed it and continued to challenge me throughout my graduate career with the difficult and often overlooked questions: "Why did you choose to do that experiment? Why does that particular result matter?". Dave, the scientific input and technical insight I received from you was paramount to the completion of my PhD. I appreciate your continued support at scientific meetings and your genuine interest in my progress throughout my graduate career, even if you are a Blue Devil. Jean, I admire the confidence that you carry into every room you enter and the confidence that you have

instilled in me to be a strong woman in science. Our conversations at CSH about managing scientific success in a healthy balance with life truly shaped the way I view my career and my career goals. I cannot describe the extent to which you have served as my role model during graduate school, and, for that, I am so incredibly thankful. Dan, when I came to UNC, I knew I wanted my chair to know the ins and outs of my personality and work ethic. I did not expect to gain a mentor and colleague who I now view as a second PI. Thank you for all the small, but incredibly significant, things you did and said to me to ensure that I made it through my PhD better than how I started. Also, a huge thank you to Piper. You got me through one of the toughest weeks of my life without even knowing it.

To the members of the Blue Pod, past and present, thank you for being the ultimate support system. Spencer and Chris, thank you for bringing an insane amount of intelligence and kindness to the workplace. I am incredibly appreciative of the interactions we shared over the past five years. Amanda, from day one you have been the glue that has held the pod together; but personally, you have been the glue that has held me together. You are one of the most amazing people I have ever met, and your love and support for others is unmatched. Every day, your friendship has pulled me through this wild ride, and I'm so grateful to have you in my life. Also, thank you from the bottom of my heart for being a Patriots fanatic. Taylor, working with you was one of the greatest professional experiences I could ever hope for—when I fell behind, you picked me up, and when you needed a hand, you encouraged me to take the lead. I treasure our friendship, and I am looking forward to seeing all the professional and personal successes you attain. Also, thank you for Phantom, who has brought me so much joy and happiness—thank you for helping me write Chapter five and for all the years of unequivocal love you have given me. Katy, you are my rock. You are the

strongest and hardest working person I know, and I will always be in awe of the beautiful person you are. You taught me to chase my dreams and surpass my limits. I cannot wait to see the amazing feats we can, and will, accomplish together.

To my friends outside of lab, you made this graduate school experience an absolute blast. Kelsey and Ian, thank you for the nights spent screaming at the Tarheels at Hickory Tavern. We made it through the past six years together, and I am so proud of you both for landing your dream jobs. Adele, you are always there right when I need you most. Every dinner and coffee date made being a graduate student exponentially better, and I truly cherish our friendship.

To the current members of the Duronio lab. I do not have the words to thank you for your support and your friendship. Our lab has become a truly unique family, and it has been the ultimate privilege to work with each of you. Chris, Ashlesha/Sharon, Jim, and Aaron, thank you for accepting me for my authentic self—you all are truly amazing friends. Mary, you took me under your wing, welcomed me into your home, and made me feel like a part of your family. There are no words that encompass what you mean to me. I love you and Tim dearly and look forward to the mischief we will certainly get ourselves into in the years to come.

This incredible journey started and will continue with the amazing people I get to call my family. To my North Carolina and New York families, thank you for your unending love and support. Grandma, you were a big reason why I came to UNC for graduate school. I am blessed to have had the opportunity to experience the last five and a half years of your life with you; this was such an amazing gift, and I love you so much. Erin, Ethan, and our little bundle of joy, Lily, you continue to enhance the quality of my life every day, and I love you

all with my entire heart. I am overjoyed to be heading back up north to spend the next years of my life with you. Mom and Dad, you epitomize the qualities of amazing parents. You have always put Erin and me first, and this body of work and this major accomplishment in my life is all for you. I love you forever and for always. We did it!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ac	Acetyl
ARS	Autonomously replicating sequence
BAC	Bacterial Artificial Chromosome
bp	Base pair
BSA	Bovine serum albumin
CDK	Cyclin-dependent kinase
CDS	Coding sequence
Cdt1	Cdc10-dependent transcript 1 gene
Cdc6	Cell division cycle 6 gene
ChIP	Chromatin immunoprecipitation
Chr	Chromosome
CMG	The replicative helicase containing Cdc45, MCM2-7. And GINS
CTR	Constant timing region
CUT&RUN	Cleavage under targets and release using nuclease
DAPI	4’,6-Diamidino-2-phenylindole
DDK	Dbf4-dependent kinase
DNA	Deoxyribonucleic acid
EdU	5-Ethynyl-2′-deoxyuridine
FACS	Fluorescence activated cell sorting
FISH	Fluorescence *in situ* hybridization
FAIRE	Formaldehyde assisted isolation of regulatory elements
G	Guanine

GFP	Green fluorescent protein
gRNA	Guide ribonucleic acid
H2Av	Histone 2a variant v
H3	Histone 3
H4	Histone 4
$H3.3^{K9R}$	Lysine to arginine mutation at the ninth lysine of histone H3.3
$H3.3^{WT}$	Wild type histone H3.3
H3K9	Ninth lysine of histone H3
H3K9R	Lysine to arginine mutation at the ninth lysine of histone H3
H4K16R	Lysine to arginine mutation at the sixteenth lysine of histone H4
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His4r	Histone H4 replacement
HisC	$Df(2L)HisC^{ED1429}$
HP1	Heterochromatin Protein 1
HWT	Histone wild-type
K	Lysine
kb	Kilobase pair
Mb	Megabase pair
MCM	Mini-chromosome maintenance
me	Methyl
MNase	Micrococcal nuclease
modENCODE	Encyclopedia of DNA elements for model organisms
nanos	Promoter for the *nanos* gene

NEB	Nuclear extraction buffer
ORC	Origin recognition complex
PBS	Phosphate buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PTM	Post-translational modification
RFP	Red fluorescent protein
RNA	Ribonucleic acid
RPKM	Reads per kilobase of transcript per million mapped reads
RPM	Reads per million
RT	Replication timing
SMW	Second mitotic wave
SuUR	Suppressor of under-replication
SSC	Saline sodium citrate
TAD	Topologically associated domain
TPM	Transcripts per million
TTR	Timing transition region
twi	Promoter for the *twist* gene
YFP	Yellow fluorescent protein
UAS	Upstream activating sequence
UTR	Untranslated region

CHAPTER 1 – INTRODUCTION

DNA replication of metazoan genomes

In an average human lifetime, 10^{16} cell divisions occur within the body (Milo et al. 2010). During each cell division, the genetic information encoded from one parent cell is passed to two daughter cells. Genome integrity is maintained during cell division through the coordination of many tightly controlled mechanisms that ensure complete and accurate genome duplication while preventing deleterious mutation and chromosome mis-segregation. If these mechanisms go awry, severe developmental outcomes, such as the onset of disease, can occur. Therefore, understanding how DNA replication is regulated in space and time is critical to understanding the fundamental aspects of cell division and disease.

Proliferating animal cells are faced with the difficult task of duplicating a large, complex genome in a short period of time during each cell division cycle. In mammals, this is accomplished through replication of the genome in temporally and spatially separated segments of approximately 400-800 kilobases, termed replication domains (Pope et al. 2014). DNA replication initiates from thousands of sites across the genome each S phase, termed origins of replication, to replicate large genomes in a timely manner. To ensure that the genome replicates once and only once per S phase, origin licensing and origin activation occur in two distinct cell cycle phases. Origins of replication are licensed during G1 phase of the cell division cycle through the concerted activities of the origin recognition complex (ORC), Cdc6, Cdt1, and the mini-chromosome maintenance (MCM) complex (Bell and

Stillman 1992; Liang et al. 1995; Coleman et al. 1996; Rowles et al. 1996; Nishitani et al. 2000). The first step in origin licensing involves ORC binding to chromatin (Duzdevich et al. 2015). Next, Cdc6 binds to ORC, which is necessary to recruit a single Cdt1-bound MCM2-7 hexamer (Duzdevich et al. 2015). The origin of replication is considered licensed when a second Cdt1-bound MCM2-7 hexamer is loaded adjacent to the first to form a head-to-head double hexamer (Evrin et al. 2009; Remus et al. 2009).

Upon S phase entry, licensed origins of replication become competent to activate, primarily through the activity of two major kinases, cyclin-dependent kinases (CDKs) and Dbf4-dependent kinases (DDKs). CDKs and DDKs phosphorylate a host of replication initiation factors, stimulating the recruitment of proteins necessary to form, and initiate bidirectional DNA synthesis from, the mature Cdc45/MCM2-7/GINS (CMG) replicative helicase. Importantly, while tens of thousands of origins are licensed in mammalian genomes each G1 phase, only a small fraction successfully activate DNA synthesis during S phase. This regulation is controlled in part by a subset of replication initiation factors (Sld2, Sld3, Dbf4, and Dpb11) that exist at limiting concentrations, providing a temporal order in which replication origins activate (Mantiero et al. 2011; Collart et al. 2013). Because origin activation is a stochastic process, only a small fraction of origins initiate synthesis during S phase, possibly due to differential activation efficiency of individual origins (Rhind et al. 2010). Although we now know the 42 protein factors required for origin activation (Yeeles et al. 2015), many unanswered questions remain regarding the mechanisms that dictate 1) where origins of replication are licensed in G1, 2) how origins of replication are selected for activation during S phase, and 3) when an origin activates DNA replication during S phase.

Regulation of DNA replication timing

Initiation of bidirectional DNA synthesis from origins of replication is staggered in space and time during S phase. This results in spatially separated regions of actively replicating DNA; this asynchrony of replication is termed the DNA replication timing (RT) program (Taylor 1958; Taylor 1960; Woodfine et al. 2004; Ryba et al. 2010; Pope et al. 2014). Although asynchronous origin firing is evolutionarily conserved among eukaryotes, the biological function and control mechanisms of RT programs are not completely understood. Importantly, altered RT of cancer-related genes correlates with changes in gene expression and contributes to malignant states (De and Michor 2011; Koren et al. 2012; Black et al. 2013; Fritz et al. 2013; Sima and Gilbert 2014; Polak et al. 2015; Rivera-Mulia and Gilbert 2016). Furthermore, genome-wide RT changes observed in cancer cells have been postulated to occur early during disease progression and may be sufficient to predict common disease-associated translocations (Koren et al. 2012; Donley and Thayer 2013; Rivera-Mulia et al. 2017; Du et al. 2019). We still do not understand whether RT change is a cause or consequence of disease progression, necessitating further understanding of the mechanisms regulating RT programs in both normal and disease contexts.

Local chromatin structure

One proposed mechanism for the regulation of RT is chromatin structure. Genome replication occurs in the context of chromatin, which is comprised of structures termed nucleosomes where ~147 bp of DNA wraps around an octamer of histone proteins. Two polypeptides each of histones H2A, H2B, H3, and H4 make up the protein octamer core of the nucleosome, and histone H1 serves as the linker histone. Histone post-translational modification (PTM) of the unstructured N-terminal histone tails modulates nucleosome

function and density along DNA, establishing chromatin states that are either “open” (euchromatin, enriched in H3/H4 acetylation (ac) and H3K4 methylation (me)) or “closed” (heterochromatin, enriched in H3K9me2/3, H4K20me3, and H3K27me3). The chromatin landscape influences the binding of *trans*-acting factors primarily through either direct binding of factors to specific histone PTMs or through recruitment of factors to open, accessible chromatin. In metazoans, chromatin is thought to influence both recruitment of abundant *trans*-acting factors (ORC, Cdc6, Cdt1 and MCM2-7) that license origins in G1 and accessibility of DNA to limiting replication initiation factors (Sld2, Sld3, Dbp11, and Dbf4) that activate origins of replication in S phase (Mantiero et al. 2011; Collart et al. 2013; Das et al. 2015; Miotto et al. 2016). This is in contrast to *S. cerevisiae*, where origins of replication are sequence defined such that ORC binds to a conserved motif termed the autonomously replicating sequence (ARS) (Stinchcomb et al. 1979). As all other eukaryotes studied to date lack sequence-defined origins of replication, the exact mechanisms through which chromatin dictates the genome-wide landscape of licensed origins remain unclear.

ORC preferentially binds to G-rich, accessible chromatin, resulting in more licensed origins within “open” euchromatic regions of the genome relative to “closed” heterochromatic regions (Delgado et al. 1998; MacAlpine et al. 2010; Cayrou et al. 2011). Furthermore, current models suggest that increased chromatin accessibility promotes the loading of more MCM complexes per ORC, further increasing the density of origins in euchromatic regions relative to heterochromatic regions (Das et al. 2015). Upon entry into S phase, replication initiation is thought to follow a stochastic model based largely on chromatin accessibility (Yang et al. 2010; Comoglio and Paro 2014; Gindin et al. 2014; Das et al. 2015). Replication initiation factors are recruited to euchromatic origins with a higher

probability than heterochromatic origins due to an increased density of licensed origins in accessible regions of the genome relative to inaccessible regions. Differential origin activation within euchromatic and heterochromatic regions of the genome contributes to the relatively earlier RT of euchromatin relative to heterochromatin (Mantiero et al. 2011; Collart et al. 2013; Das et al. 2015). However, correlations between early RT and euchromatin and late RT and heterochromatin are not absolute, suggesting additional modes of regulation beyond chromatin structure.

Transcriptional activity

Transcriptionally active regions of the genome tend to replicate earlier during S phase whereas transcriptionally repressive regions tend to replicate later during S phase (Goldman et al. 1984; Lubelsky et al. 2014). Because transcriptional activity is also strongly correlated with chromatin structure, it remains unclear whether the relationship between transcriptional activity and RT is a consequence of chromatin structure or if transcriptional activity directly influences RT. Despite strong correlations between active transcription and early RT, extremely high levels of transcription have been shown to inhibit replication initiation (Martin et al. 2011). Furthermore, transcriptional activity can displace the MCM complex, changing the origin location from the initial site of ORC binding (Gros et al. 2015; Powell et al. 2015). Although transcriptional activity can directly influence origins of replication, increasing evidence suggests that RT and transcriptional activity are regulated by a common chromatin environment (Lubelsky et al. 2014).

Transcriptionally active, euchromatic regions of the genome contain more origins of replication than lowly transcribed, heterochromatic regions of the genome (MacAlpine et al. 2010; Eaton et al. 2011; Miotto et al. 2016). Enrichment of active replication origins at

promoters is thought to be driven by the open chromatin conformation of promoters—active transcription maintains accessibility at promoters, thus promoting origin licensing in G1 or activation in S phase (MacAlpine et al. 2010; Dellino et al. 2013; Miotto et al. 2016). Dellino et al. mapped ~13,000 ORC1 binding sites and found that almost all mapped origins were associated with transcription start sites of either coding or noncoding RNAs (Dellino et al. 2013). Similarly, Miotto et al. mapped ORC2 binding genome-wide and found similar binding profiles to the independently derived ORC1 data (Miotto et al. 2016). Importantly, the ORC enrichment at promoters is most likely a consequence of the open chromatin environment as there is only a modest genome-wide correlation between ORC binding and transcriptional activity (Miotto et al. 2016).

Three-dimensional genome architecture

While both local chromatin structure and transcriptional activity influence genome-wide RT from yeast to humans, emerging evidence points to three-dimensional genome architecture as a previously unappreciated, key contributor to RT control (Rivera-Mulia and Gilbert 2016). The genome is organized within subnuclear compartments such that early replicating regions tend to be located at the nuclear interior (active compartment "A"), and late replicating regions tend to be located at the nuclear or nucleolar periphery (inactive compartment "B") (Visser et al. 1998; Lieberman-Aiden et al. 2009). Within each compartment, the genome is further organized into topologically associated domains (TADs), sub megabase-sized DNA sequences that display frequent physical interaction within three-dimensional space. Remarkably, replication domain boundaries share a near one-to-one correlation with topologically associated domain (TAD) boundaries (Moindrot et al. 2012; Pope et al. 2014). It is now understood that bioinformatically-defined replication domains

and TADs and cytologically-defined replication foci likely represent the same structures (Xiang et al. 2018). However, TAD boundaries are not absolutely required for maintenance of RT as TAD boundary disruption, either through deletion of DNA sequences at TAD boundaries or depletion of the protein components required for establishing interactions between adjacent TADs, has no effect on RT (Oldach and Nieduszynski 2019; Sima et al. 2019). Interestingly, while TAD boundaries are not required for RT maintenance, emerging evidence suggests that sequence elements within TADs (early-replication control elements, ERCEs) drive interactions between TADs required for maintenance of early RT (Sima et al. 2019). While further study is necessary to provide mechanistic insight into the relationship between three-dimensional genome architecture and RT, our understanding of RT control mechanisms is beginning to parse out correlative versus causal relationships.

Cell lineage

During animal development, cells undergo progressive changes in genome structure and function in order to generate more differentiated cell types. Transcriptional programs differ between cell types, and cell type-specific transcriptomes are reflected by genome-wide changes in both three-dimensional arrangement of DNA within the nucleus and local chromatin structure. Interestingly, replication domain boundaries and TAD boundaries are stable structural units during cellular differentiation where the RT and subnuclear localization of individual replication domains/TADs differs in a lineage-specific manner (Pope et al. 2014). Approximately 50% of the genome displays differential RT between cell types, where constant timing regions (CTRs) display unchanged RT during development and timing transition regions (TTRs) progressively advance RT during development (Hiratani et al. 2008). The subnuclear position of TTRs changes coincidentally with RT change—the

TTR is positioned in compartment A in cells where RT has advanced. In fact, Heinz et al. demonstrated that manipulating the nuclear position of pericentric heterochromatin was sufficient to advance its RT in mammalian cells (Heinz et al. 2018).

Cellular differentiation provides a unique system to track the dynamics of RT, transcription, chromatin accessibility, and three-dimensional genome architecture in developmental time. Studying RT in this fashion has revealed multiple instances where genome architecture and RT are mechanistically separable. Studies in the early *Drosophila* embryo have demonstrated that the onset of late replication precedes the establishment of constitutive heterochromatin suggesting that the hallmarks of heterochromatin (H3K9me2/3 and HP1a enrichment) are not required for late RT (Yuan and O'Farrell 2016). Furthermore, establishment of RT programs was shown to anticipate transcriptional programs in the early zebrafish embryo (Siefert et al. 2017) while transcriptional change often preceded RT change during differentiation of mammalian cells (Rivera-Mulia et al. 2015). From studies conducted in mammalian cells, we now know that correlations between RT, transcription, chromatin, and three-dimensional genome architecture become stronger as cells differentiate. Interestingly, strong correlations were shown to be restricted to genes located in CTRs (Rivera-Mulia et al. 2015), and these correlations are much weaker in TTRs (Besnard et al. 2012; Takebayashi et al. 2012; Dileep et al. 2015). Collectively, these data raise the possibility that RT may be regulated differently between CTRs and TTRs and further emphasize that our current understanding of RT control in metazoan species is incomplete.

Replication timing: Genome instability and human disease

Normal RT programs contribute to genome instability

Many biological processes contribute to genomic instability, including the normal RT program. RT has been proposed to contribute to the non-random genome-wide distribution of mutations, where early replicating regions are more susceptible to *trans* chromosomal rearrangements and late replicating regions are more prone to *cis* translocations and to point mutations (Watanabe et al. 2002; Stamatoyannopoulos et al. 2009; Cui et al. 2012; Sima and Gilbert 2014; Supek and Lehner 2015; Du et al. 2019). Furthermore, late replicating regions and origin-depleted regions that are passively replicated by an adjacent origin (TTRs), experience a greater overall mutational burden than early replicating regions (Watanabe et al. 2002; Hiratani et al. 2008; Watanabe and Maekawa 2010; De and Michor 2011). Many common structural mutations, fragile sites, hotspots for copy number alterations, and genomic rearrangement sites in cancer are found in TTRs, possibly due to the complicated nature of replicating these origin-depleted regions (Watanabe et al. 2002; Donley and Thayer 2013; Rhind and Gilbert 2013). It has been postulated that minimizing the mutational burden in early replicating regions of the genome helps to prevent mutation of ubiquitously expressed "housekeeping" genes. Consequently, tissue-specific genes located in TTRs and late replicating gene-poor regions of the genome experience the bulk of the mutational burden.

RT alterations in cancer

Recent single cell RT profiling studies demonstrate that RT programs are highly stable between individual cells of the same cell type, whereas properties such as epigenetic marks and transcription show dynamic cell-to-cell heterogeneity (Dileep and Gilbert 2018;

Ozgyin et al. 2019; Takahashi et al. 2019). The robustness of RT programs is further highlighted by the fact that almost every attempt to disrupt RT, including genetic perturbation of key RT control factors, results in little to no effect on genome-wide RT (Yokochi et al. 2009; Pope et al. 2014; Foti et al. 2016; Armstrong et al. 2018; Oldach and Nieduszynski 2019; Sima et al. 2019). Despite the robust nature of RT, RT programs are commonly altered in cancer, and altered RT programs have been proposed to be an early epigenetic event in disease progression (Ryba et al. 2012; Koren et al. 2014; Rivera-Mulia et al. 2017). Importantly, the proportion of, and the specific loci within, the genome that display altered RT are cancer type-specific with, for example, LNCaP prostate cancer cells displaying altered RT at 5.7% of the genome and acute lymphoblastic leukemia patient cells displaying altered RT at 9-18% of the genome (Ryba et al. 2012; Du et al. 2019). Because, in some instances, RT can differentiate disease and normal tissue in ways conventional transcriptomic analysis cannot, RT profiling has been proposed as a potential diagnostic tool for cancers with unique RT signatures (Rivera-Mulia et al. 2017).

Amongst the classes of RT alterations observed in cancer is asynchronous replication, or the differential replication timing of homologous loci that normally replicate at the same time during S phase (Amiel et al. 1998; Litmanovitch et al. 1998; Korenstein-Ilan et al. 2002). Interestingly, asynchronous replication has not only been observed for cancer-related genes in cancer cells, but has also been observed in peripheral lymphocytes of patients with solid tumors, pre-malignant cells, and in cells of individuals with a predisposition to cancer, suggesting that asynchronous replication may be an early event in cancer progression (Amiel et al. 1998; Reish et al. 2003; Cytron et al. 2011). Furthermore, in cancer cells, the RT of at least one entire chromosome is commonly severely delayed, and its replication may persist

into G2 phase or even into mitosis (Smith et al. 2001; Donley and Thayer 2013; Platt et al. 2018). A chromosome-wide delay in RT is unsurprisingly associated with highly aneuploid karyotypes and a 30 to 80-fold increase in chromosomal rearrangement of the affected chromosome (Smith et al. 2001; Breger et al. 2004; Breger et al. 2005). Altogether, RT programs influence the non-random mutation distribution within cells, contribute to the early events of disease progression, and propagate genome instability of cancer cells emphasizing the need for a better understanding of the molecular mechanisms regulating RT programs in both normal and disease contexts.

CHAPTER 2- CHROMATIN CONFORMATION AND TRANSCRIPTIONAL ACTIVITY ARE PERMISSIVE REGULATORS OF DNA REPLICATION INITIATION IN *DROSOPHILA*[1]

Introduction

Animal cells duplicate large, complex genomes by initiating replication at distinct locations within the genome at different times during S phase. An evolutionarily conserved feature of this regulatory paradigm is a temporal order of DNA replication initiation events that results in characteristically early and late replicating regions of the genome (Rhind and Gilbert 2013). Such "replication timing" (RT) programs appear at early stages of animal development and ensure genome integrity during cell proliferation (Shermoen et al. 2010; Mantiero et al. 2011; Collart et al. 2013; Hamperl and Cimprich 2016; Yuan and O'Farrell 2016; Almeida et al. 2018). Importantly, RT is associated with mutational burden and SNP density, as spontaneous mutations occur less frequently in early compared to late replicating regions of the genome (Stamatoyannopoulos et al. 2009; Donley and Thayer 2013). Furthermore, perturbed RT is thought to be an early epigenetic event that predisposes cancer and disease-associated genome rearrangement (Ryba et al. 2012; Donley and Thayer 2013). Notwithstanding their importance, mechanisms that control where and when DNA replication initiates within an animal genome remain poorly understood.

[1] This chapter previously appeared as an article in Genome Research. The original citation is as follows: Armstrong, R.L., Penke, T.J.R., Strahl, B.D., Matera, A.G., McKay, D.J., MacAlpine, D.M., and Duronio, R.J., 2018. Chromatin conformation and transcriptional activity are permissive regulators of DNA replication initiation in *Drosophila*. *Genome Research.* Accepted.

In contrast to replication initiation in single-celled eukaryotes such as budding yeast, replication of animal genomes does not initiate at well-defined sequence motifs (Bell and Stillman 1992; MacAlpine et al. 2010; Miotto et al. 2016). Rather, two levels of genome organization have emerged as putative regulators of replication initiation: three dimensional arrangement of DNA within the nucleus and local chromatin structure, characterized in part by differential DNA accessibility (i.e. differential nucleosome occupancy) (Hiratani et al. 2008; Pope et al. 2014; Heinz et al. 2018). Current models posit that these features of genome organization regulate replication by influencing *trans*-acting factor recruitment to sites of replication initiation (i.e. origins) (Mantiero et al. 2011; Collart et al. 2013; Pope et al. 2014; Das et al. 2015; Miotto et al. 2016; Rivera-Mulia and Gilbert 2016). In all metazoan organisms examined to date, transcriptionally active, accessible euchromatin generally replicates early during S phase, whereas transcriptionally repressive, inaccessible heterochromatin generally replicates late (Bell et al. 2010; Eaton et al. 2011; Lubelsky et al. 2014). Despite strong genome-wide correlations between replication and chromatin structure in animal cells, efforts to determine a causal relationship between the two have been hampered by imprecise methods for manipulating chromatin structure *in vivo*. We therefore developed an approach for altering the distribution of accessible chromatin throughout the genome and determined if and how these changes in chromatin structure affect genome replication.

Strategies to manipulate chromatin structure in animal cells often involve perturbation of factors that establish, interpret, or remove histone post-translational modifications (PTMs). Although informative, these studies cannot precisely determine functional roles for histone PTMs in DNA replication because most histone-modifying enzymes also have non-histone

substrates that may participate in DNA replication (Glozak et al. 2005; Huang and Berger 2008). Therefore, to reduce potential pleiotropic effects of mutating histone-modifying enzymes, we employed a strategy in *Drosophila* to more precisely manipulate chromatin structure by mutating the histone genes themselves, an approach that is not currently feasible in other animal models. This strategy involves deleting the endogenous wild type histone genes and replacing them with transgenic copies encoding a single amino acid substitution that prevents PTMs of a particular histone residue (Günesdogan et al. 2010; McKay et al. 2015). Here, we determine how two different histone mutations that affect chromatin organization and transcription in heterochromatin (*H3K9R*) and euchromatin (*H4K16R*), respectively, affect DNA replication initiation throughout the genome.

Materials and Methods

Complete genotypes

"12x*HWT*" (Histone Wild Type) refers to a control Bac-based transgene containing 12 copies of the 5kb histone wild type repeat unit containing all five replication dependent histone genes (McKay et al. 2015). "12x*H3K9R*" and "12x*H4K16R*" are identical transgenes except with a Lys to Arg substitution mutation at the 9th residue of histone H3 and 16th residue of histone H4, respectively.

HWT: *yw; ΔHisC, twi-Gal4/ΔHisC,UAS-2xEYFP; 12xHWT/+* (McKay et al. 2015)

H3K9R**:** *yw; ΔHisC, twi-Gal4/ΔHisC,UAS-2xEYFP; 12xH3K9R/+* (Penke et al. 2016)

H4K16R**:** *yw; ΔHisC, twi-Gal4/ΔHisC,UAS-2xEYFP; 12xH4K16R/+*

H3.3^{WT}; H3^{WT}**:** *yw; H3.3A^{2x1}, ΔHisC, twi-Gal4/Df(2L)BSC110, ΔHisC,UAS-2xEYFP; 12xHWT/+* (Penke et al. 2018)

***H3.3^{K9R}; H3^{K9R}*:** *yw, H3.3B^{K9R}; H3.3A^{2x1}, ΔHisC, twi-Gal4/Df(2L)BSC110, ΔHisC,UAS-2xEYFP; 12xH3K9R/+* (Penke et al. 2018)

Zygotic, replication-dependent *HWT*: *yw; ΔHisC, twi-Gal4/ΔHisC,UAS-2xEYFP; 12xHWT/+*

Zygotic replication-dependent *H4K16R*: *yw; ΔHisC, twi-Gal4/ΔHisC,UAS-2xEYFP; 12xH4K16R/+*

Maternal/zygotic, replication-dependent *HWT*: *yw; ΔHisC, twi-Gal4/ΔHisC,UAS-2xEYFP; 12xHWT/+* (from mothers of genotype *ΔHisC, UAS-2xeYFP; 12xHWT*)

Maternal/zygotic, replication-dependent *H4K16R*: *yw; ΔHisC, twi-Gal4/ΔHisC,UAS-2xEYFP; 12xH4K16R/+* (from mothers of genotype *ΔHisC, UAS-2xEYFP; 12xH4K16R*)

Zygotic, replication-dependent and replication-independent *HWT*: *yw; ΔHisC, twi-Gal4/ΔHisC,UAS-2xEYFP; 12xHWT, His4$^{15-4}$/His4r$^{15-4}$*

Zygotic replication-dependent and replication-independent *H4K16R*: *yw; ΔHisC, twi-Gal4/ΔHisC,UAS-2xEYFP; 12xH4K16R, His4r$^{15-4}$/His4r$^{15-4}$*

Generation of *H3K9R* and *H4K16R* mutant genotypes

All fly stocks were maintained on standard corn medium and crossing schemes to generate replication-dependent histone genotypes were performed as in (Penke et al. 2016). For first instar larval brain EdU experiments, the following crosses were performed: ***H3.3^{WT}; H3^{WT}***) *yw; H3.3A^{2x1}, ΔHisC, twi-Gal4/CyO* females were crossed to *yw; Df(2L)BSC110, ΔHisC,UAS-2xEYFP/CyO; 12xHWT* males and for ***H3.3^{K9R}; H3K9R***) *yw*, *H3.3^{K9R}; H3.3A^{2x1}, ΔHisC, twi-Gal4/CyO* females were crossed to *H3.3^{K9R}; Df(2L)BSC110, ΔHisC,UAS-2xEYFP/CyO; 12xH3K9R/+* males. Note that only animals containing either the 12x*HWT* or 12x*H3K9R* transgenes can complete embryogenesis as in Penke et al. 2018. To

generate zygotic, replication-dependent *HWT* and *H4K16R* mutants, *ΔHisC, twi-Gal4/CyO* mothers were crossed to *ΔHisC, UAS-2xEYFP/CyO; 12xHWT/12xHWT* or *ΔHisC, UAS-2xEYFP/CyO; 12xH4K16R/12xH4K16R* fathers, respectively. To generate flies where both the maternal and zygotic contribution of histones were *HWT* or *H4K16R* mutant, *ΔHisC, UAS-2xeYFP; 12xHWT* or *ΔHisC, UAS-2xEYFP; 12xH4K16R* mothers, respectively, were crossed to *ΔHisC, twiGal4/CyO* fathers. To generate zygotic, replication-dependent and replication-independent *HWT* and *H4K16R* mutants, *ΔHisC, twi-Gal4/CyO; $His4r^{15-4}$* mothers were crossed to *ΔHisC, UAS-2xEYFP/CyO; 12xHWT, $His4r^{15-4}$* or *ΔHisC, UAS-2xEYFP/CyO; 12xH4K16R, $His4r^{15-4}$* fathers, respectively. For each *H4K16R* viability experiment, groups of fifty GFP+ first instar larvae of each genotype were separated from their wild type siblings into vials of standard corn medium and allowed to complete development.

CRISPR-Cas9 Mutagenesis of *His4r*

Two different gRNA oligos targeting the 5'UTR (target sequence: 5'-CCTGTCAAATGAACGTTTACCTT-3') and the 3' adjacent intergenic region (target sequence: 5'- CCGAAAATAAGGTCCAACAAACT-3') of the *His4r* gene were inserted into the pCFD3 vector. gRNA constructs were co-injected into embryos expressing Cas9 from the *nanos* promoter (*nanos-cas9*) (Kondo and Ueda 2013). A 721-bp deletion spanning the entire *His4r* CDS (*$His4r^{15-4}$*) was identified by PCR and is referred to as *His4rΔ* in this manuscript. The *His4rΔ* allele removes sequence between the following 5' and 3' 20nt flanking sequences, respectively: 5'-CTTATTAACAACAGTTTTCA-3' and 5'-CGCTTAGGGAGCACACAAAT-3'.

Culture conditions for embryo sorting

A Union Biometrica BioSorter for large particle flow cytometry equipped with a 488-nm solid state laser and accompanying FlowPilot software was used for identification and high throughput isolation of GFP-positive Δ*HisC, UAS-2xEYFP/*Δ*HisC, twi-GAL4* mutant embryos from their GFP-negative siblings. For this purpose, three hundred to four hundred *ΔHisC, twi-Gal4/CyO* females and 100 *ΔHisC, UAS-2xEYFP/CyO; 12xHWT/12xHWT*, 100 *ΔHisC,UAS-2xEYFP/CyO;* 12x*H3K9R/12xH3K9R*, or 100 *ΔHisC,UAS-2xEYFP/CyO;* 12x*H4K16R/12xH4K16R* males were placed in a large embryo collection cage (fits 100mm petri dishes) at 25°C and allowed to lay eggs on apple juice agar plates. Overnight collections were dechorionated in 100% bleach for two minutes and collected in embryo wash buffer (0.7% NaCl, 0.07% Triton X-100) prior to embryo sorting. Aliquots of one hundred GFP-positive embryos were transferred to vials containing standard corn medium and cultured at 25°C to obtain third instar larvae.

Sample preparation for FACS and sequencing

Third instar wing imaginal discs were dissected over a period of four hours and stored in Grace's insect medium (supplemented with L-Glutamine, 3.33g/L Lactalbmin Hydrolysate, and 3.33g/L Yeastolate) on ice prior to nuclear isolation. Nuclear isolation was performed similarly to (Ma and Weake 2014) with the following adjustments. In brief, a 2mL dounce homogenizer was pretreated with nuclear extraction buffer (NEB; 10mM HEPES-KOH, pH 7.5; 2.5 mM $MgCl_2$; 10mM KCl) supplemented with 5% BSA and placed on ice. Subsequently, wing imaginal discs were transferred to the empty dounce and incubated in NEB for five minutes on ice, disrupted twenty times with the loose pestle, incubated for ten minutes on ice, and disrupted 15 times with the tight pestle. Homogenate was filtered over a

CellTrics 30 μm filter and isolated nuclei were stained with 1.5μg/mL DAPI prior to FACS. Nuclei were sorted into G1, S, and G2 populations based on DNA content as measured by DAPI intensity on a FACSAria II or III (using NEB-0.1% Tween sheath). Gates were chosen conservatively to prevent contamination of either fraction with nuclei from a neighboring fraction. Analyses of cell cycle indices were performed on DAPI profiles generated from FACS using the Dean-Jett Fox model included with the FlowJo software (Dean and Jett 1974; Fox 1980). Isolated populations of nuclei were pelleted, flash frozen, and stored at -80°C prior to DNA isolation and library preparation. Genomic DNA libraries were prepared with the Rubicon ThruPLEX DNA-seq kit and sequencing was performed on an Illumina HiSeq 2500 in the UNC-Chapel Hill High Throughput Sequencing Facility.

Sequence data analysis

Analyses were performed using R (Team 2017).

Replication Timing Profiles

Single-end 50-bp reads from G1, S, and G2 samples were aligned to the dm6 reference genome (Release 6.04) using Bowtie 2 (v2.3.2) default parameters (Langmead et al. 2009). Two S phase replicates and one G1 replicate were generated for each genotype. Reads with a MAPQ score greater than 10 were retained using SAMtools (v1.6) (Li et al. 2009). BEDTools coverage (v2.25.0) was used to determine the number of reads mapping to 100kb windows across the genome (10kb slide) (Quinlan and Hall 2010). Read counts at each 100kb window were normalized to read depth (reads per million; RPM). Other window sizes were processed similarly. To generate a replication timing (RT) value for a particular window, the RPM value of each S phase replicate was divided by the RPM G1 value and the resulting quotient was $\log_2$ transformed. The $\log_2$ S/G1 quotients for each duplicate sample

were then averaged. As an additional control, another RT value was generated using half the G2 value, which should be equivalent to the G1 value. RT profiles were generated by plotting the RT value at each window versus the genomic location. RT profiles normalized to G1 or G2 copy number controls were very similar; we therefore used G1 for all subsequent analyses. LOESS regression lines using loess.model were created to smooth RT profiles (span=0.02 for chromosome arms, span=0.05 for Chromosome 4). Although the genomic location where RT regression lines changed direction was similar across all genotypes and replicates, we note that the range of RT values for *HWT* female samples was slightly smaller than the other two genotypes. *HWT* female samples exhibited a higher percentage of cells in S phase, which decreased the clarity of the G1/S boundary when performing FACS. We speculate that a small increase in the number of late G1 cells in S phase populations limited the dynamic range of *HWT* female RT values. We therefore used quantile normalization through the preprocess Core R package to equalize the dynamic range of RT values for each female genotype (Bolstad 2016). We note that regions altered in *H3K9R* mutants compared to *HWT* were similar without quantile normalization. The limma statistical package was used to identify windows with significantly altered RT values between *HWT* and *H3K9R* female, *HWT* and *H4K16R* female, and *HWT* and *H4K16R* male samples (lmFit, adjusted p value Benjamini and Hochberg, $p<0.01$) (Newville et al. 2014). The adjusted p value corrects for multiple testing. An additional significance parameter of an absolute $\log_2$ fold-change greater than 0.1 was included to increase stringency of the significance threshold obtained from limma. Similar results were obtained using reads not filtered by MAPQ score. Differences in RT profiles were not due to differences in G1 samples (Fig S1). Coordinates of various

chromatin states were obtained from (Kharchenko et al. 2011) and converted to dm6 coordinates using the UCSC liftOver tool (Karolchik et al. 2004).

Wild-type Replication Timing Characterization

To calculate replication domain sizes, we identified the genomic coordinates halfway between each peak and valley of an RT profile and determined the distance from one halfway point to the next. We used modENCODE ChIP-seq data from whole 3rd instar larvae to calculate histone PTM enrichment at 100kb windows across the genome (ftp://data.modencode.org/D.melanogaster/Histone-Modification/ChIP-seq/raw-seqfile_fastq/). Accession numbers for each data set are as follows: H3K36me3 (GSE47248), H3K4me1 (GSE47282), H3K4me2 (GSE47261), H3K4me3 (GSE49491), H3K79me1 (GSE49492), H3K27ac (GSE49488), H3K79me2 (GSE49493), H3K79me3 (GSE49494), H4K20me1 (GSE47254), H2Bubi (GSE49487), H3K36me1 (GSE47249), H3K23ac (GSE47257), H3K9ac (GSE48510), H3K9me1 (GSE47289), H3K9me2 (GSE47260), H3K9me3 (GSE47258), H3K9acS10P (GSE47288), H2Av (GSE47259), and H4K16ac (GSE49497) (Roy et al. 2010). For each histone PTM, raw reads for two ChIP replicates and two input replicates were aligned to the genome using Bowtie 2 (v2.3.2) (Langmead et al. 2009). BEDTools coverage (v2.25.0) was used to count the number of reads mapping to each 100kb window, and the resulting counts were normalized to read depth (Quinlan and Hall 2010). Histone PTM enrichment for each replicate was calculated by dividing the ChIP normalized read counts by the input for each replicate; the resulting values were then averaged. All windows were ordered by RT value and split into five equally sized categories (early, early/mid, mid, mid/late, and late). Average PTM enrichment values of all windows in a category were calculated and represented as a heatmap using the R package pheatmap.

RNA-seq data from 3rd instar imaginal wing discs (from GSE85374) (Penke et al. 2016) was used to calculate transcript density or transcript activity at 100kb windows. The imaginal wing disc transcriptome was assembled using Cufflinks (v2.2.1) (Trapnell et al. 2012) with the following parameters: library-type fr-firsttrand, masked rRNA, and provided dm6 transcriptome obtained from Flybase release 6.04. Subsequently, the number of transcripts overlapping each 100kb window was determined. To calculate transcript activity of a window, the normalized read per million of each transcript overlapping a window was summed.

FAIRE, HP1a, and RNA Analyses

For *H3K9R* experiments, RNA reads from three *HWT* and three *H3K9R* replicates were aligned using TopHat default parameters (v2.1.1) (Trapnell et al. 2012), and a transcriptome was generated using Cufflinks (v2.2.1, see above for parameters). Previously, we showed that the *H3K9R* mutation causes widespread de-repression of transposons (Penke et al. 2016). Therefore, in the current analysis, we combined the Cufflinks generated transcriptome with transposons annotated by RepeatMasker (Smit et al. 2013-2015). For *H4K16R* experiments, 30 wing imaginal discs per replicate were homogenized in Trizol and flash frozen in liquid nitrogen. RNA was chloroform extracted and isopropanol precipitated before column purification on Qiagen RNeasy purification with DNase digestion. Transcriptomic libraries were prepared from cDNA from total RNA (ribo-minus) with the Total RNA TruSeq Stranded Ribo Zero Gold kit. ERCC spike-in Mix 1 (ThermoFisher Scientific) was included and RNA-seq reads were normalized to the total number of reads mapping to the ERCC reference. Sequencing was performed on an Illumina HiSeq 2500 in the UNC-Chapel Hill High Throughput Sequencing Facility. Single-end 50bp RNA reads

from three *HWT* female, three *HWT* male, three *H4K16R* female, and three *H4K16R* male replicates were aligned using TopHat default parameters (v2.1.1) (Trapnell et al. 2012), and a transcriptome was generated using Cufflinks (v2.2.1, see above for parameters). For both *H3K9R* and *H4K16R* experiments, raw counts of RNA reads at each transcript were used as input for edgeR statistical analysis (p value <0.01) (Robinson et al. 2010; McCarthy et al. 2012). We then identified transcripts within or that overlapped each 10kb window and selected the transcript with the lowest p value.

To determine RT values at 10kb windows, we used the previously calculated $\log_2$ fold-change and p values from 100kb windows. We used RT values from 100kb windows as this size closely matches average replication domain size (~100-200kb), but similar results were obtained using RT values determined from 10kb windows (Fig S1). For each 10kb window, we calculated the median fold change and median p value of all overlapping 100kb windows. 10kb windows were identified as having significantly altered RT between *H4K16R* or *H3K9R* and *HWT* if $p<0.05$ (adjusted for multiple testing) and the absolute $\log_2$ fold-change was at least 0.1. To focus our analysis on more mappable regions of the genome, we analyzed 10kb windows on the major chromosome scaffolds (Chr 2L, Chr 2R, Chr 3L, Chr 3R, Chr 4, and Chr X) that, for *H3K9R* experiments, had an average FAIRE and HP1a counts per million (CPM) value of greater than zero. Comparisons of RT, FAIRE, HP1a, and RNA signal between *H3K9R* and *HWT* samples were performed with all reads or "uniquely" mapping reads (MAPQ>10) with similar results.

To calculate transposon families with significantly altered RNA levels, we summed raw counts of all individual transposons within a family and used edgeR as described above to determine significance (p value <0.05). In addition to transposon families, all transcripts

identified in Cufflinks were included in this edgeR analysis to facilitate modeling of variability. Data was visualized using the Integrative Genomics Viewer.

Immunofluorescence

Third instar wandering larvae were dissected and the carcasses inverted to expose attached imaginal discs and incubated for 60' in 0.1mg/mL EdU. Tissues were then fixed in 3.7% paraformaldehyde in PBS for 25 min. EdU incorporation was detected using the Click-It EdU Alexa Fluor 488 Imaging Kit (ThermoFisher Scientific). Carcasses were washed for 10 min in PBS-Tx (3% Triton X-100), then treated with 200ug/mL RNaseA in PBS-Tx for 2h and washed for 1 h in PBS-Tx. Individual imaginal discs were then separated from the carcass and groups of discs were successively incubated for 20 min in each of four pre-hybridization solutions: 1) 80% PBS-Tx, 20% pHM (50% formamide, 4xSSC, 100mM NaH_2PO_4 pH 7.0, 0.1% Tween 20), 2) 50% PBS-Tx, 50%pHM, 3) 20%PBS-Tx, 80% pHM, 4) 100%pHM. Denatured 359bp probe (Joyce et al. 2012) was hybridized with wing discs overnight at 37°C at 450rpm in an Eppendorf tube. Discs were successively incubated in four post-hybridization solutions for 20 min at 37°C at 800rpm: 1) 50% formamide, 2XSSC, 2) 40% formamide, 2XSSC, 3) 30% formamide, 70% PSS-Tw (1x PBS, 0.1% Tween20), 4) 20% formamide, 80% PBS-Tw and three post-hybridization solutions at 25°C: 1) 10% formamide, 90% PBS-Tw, 2) PBS-Tw, 3) PBS-Tx. DNA was stained with DAPI, and the discs were mounted in ProLong Gold antifade reagent and imaged on a Leica confocal microscope.

First instar larval brains were incubated in 0.1mg/mL EdU for 60' and subsequently fixed for 40 minutes in 3.7% paraformaldehyde. EdU was detected using the Click-It EdU Alexa Fluor 488 Imaging Kit (ThermoFisher Scientific).

Results

Profiling replication timing in a *Drosophila* tissue

To probe the relationship between chromatin structure and replication in an intact animal, we adapted a genome-wide measure of RT for use in *Drosophila* wing imaginal discs, a relatively simple epithelium of proliferating diploid cells (Koren et al. 2014; Sasaki et al. 2017; Siefert et al. 2017). Our method is based on the premise that in S phase cells early replicating DNA sequences are over-represented relative to late replicating ones, due to a higher probability of replication initiation (Rhind et al. 2010; Mantiero et al. 2011; Collart et al. 2013; Das et al. 2015). Consequently, RT data are a proxy for the propensity of replication initiation in a particular region of the genome. We performed whole-genome sequencing on DNA isolated from populations of G1 and S phase nuclei collected from wing discs by fluorescence-activated cell sorting (FACS) (Fig 2.1A). Replication profiles were generated by determining the $\log_2$ transformed S/G1 read count at 100kb intervals using a 10kb slide across the genome (Materials and Methods; Fig 2.1B; Fig 2.2, 2.3A), where larger values indicate earlier replication and smaller values indicate later replication. We chose 100kb windows with a 10kb slide because they produced the least amount of noise relative to smaller windows (Fig 2.2). RT values generated from independent S phase samples were highly reproducible (Fig 2.4).

Figure 2.1. Measuring genome-wide replication timing *in vivo*.

A) Experimental paradigm: (1) Nuclei were FACS sorted into G1 (yellow), S (red) and G2 (blue) populations based on DNA content. (2) Sequenced DNA was mapped to the dm6 genome. More reads map to early than late replicating sequences. (3) Log_2 S/G1 ratio generates RT profiles. Normalizing to G1 or G2 phase controls gave similar results. B) LOESS regression line showing average *yw* ("*yellow, white"* control genetic background used for all fly lines) RT values (log_2 S/G1) in 100kb windows with 10kb slide across Chr 2 and 3. Chromosome schematics show approximate locations of constitutive pericentric heterochromatin (green) and largely euchromatic arms (blue) (Riddle et al. 2011; Hoskins et al. 2015). C) Heatscatter plot of *yw* log_2 S/G1 (RT) versus gene density at all 10kb windows across the genome with LOESS regression line (black). D) Heatmap of relative modENCODE histone PTM enrichment in bins of equally sized RT quintiles (early, early/mid, mid, mid/late, and late) generated using RT values (log_2 S/G1) within 100kb windows. modENCODE data is from third instar larvae (Celniker et al. 2009) (see supplementary materials for accession numbers). Color indicates average enrichment of all windows within a quintile. Scale of heatmap was capped at 1.4 to better represent distribution of values as H3K9me2/me3 was greatly enriched in late replicating domains compared to other PTMs (see Fig S2E for non-capped H3K9me2/me3 heatmap). E) Plot of transposon number in 100kb windows across Chr 3R with RT quintile (as determined in D) indicated by color. Experiments were performed in collaboration with Taylor Penke.

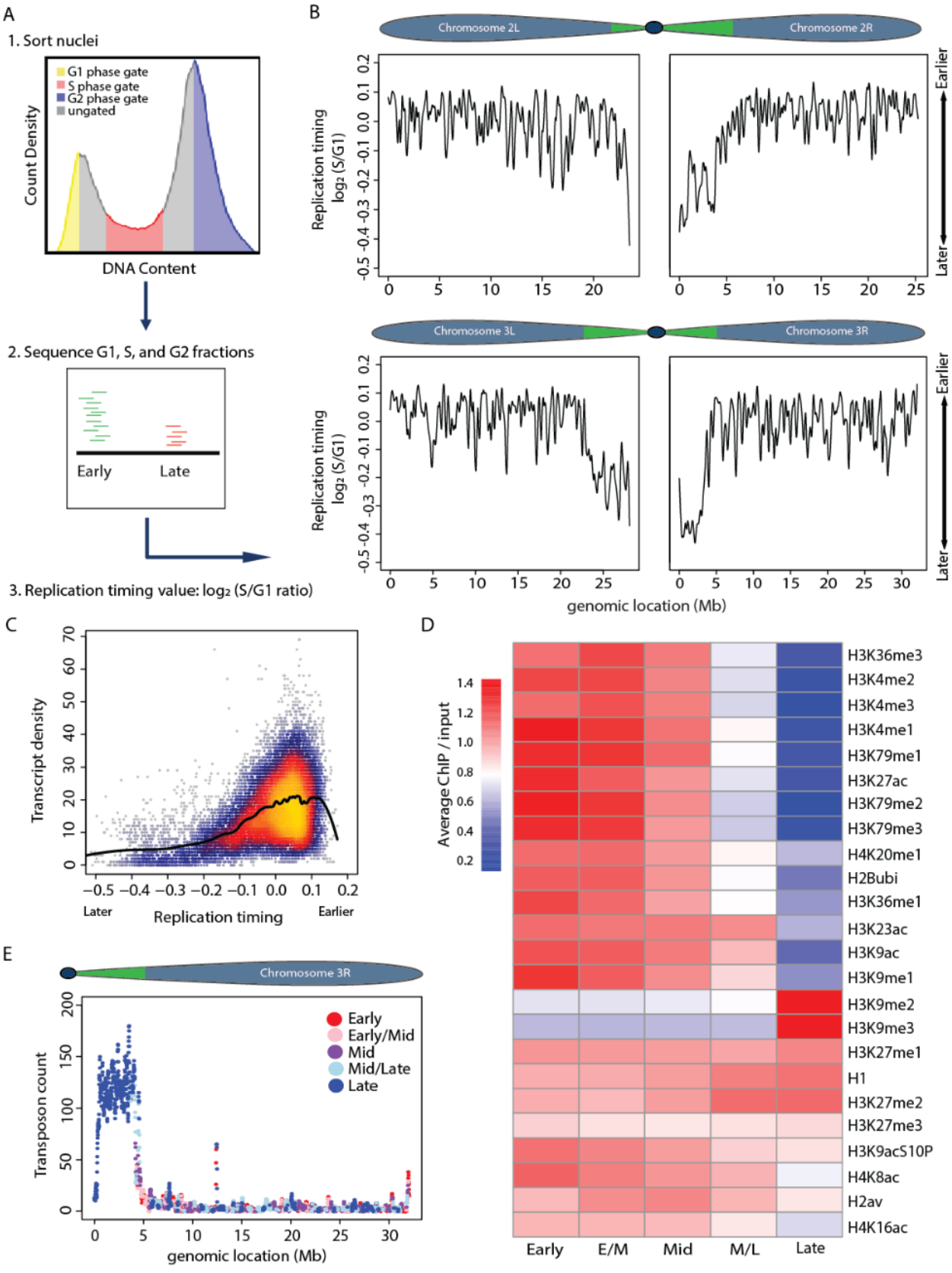
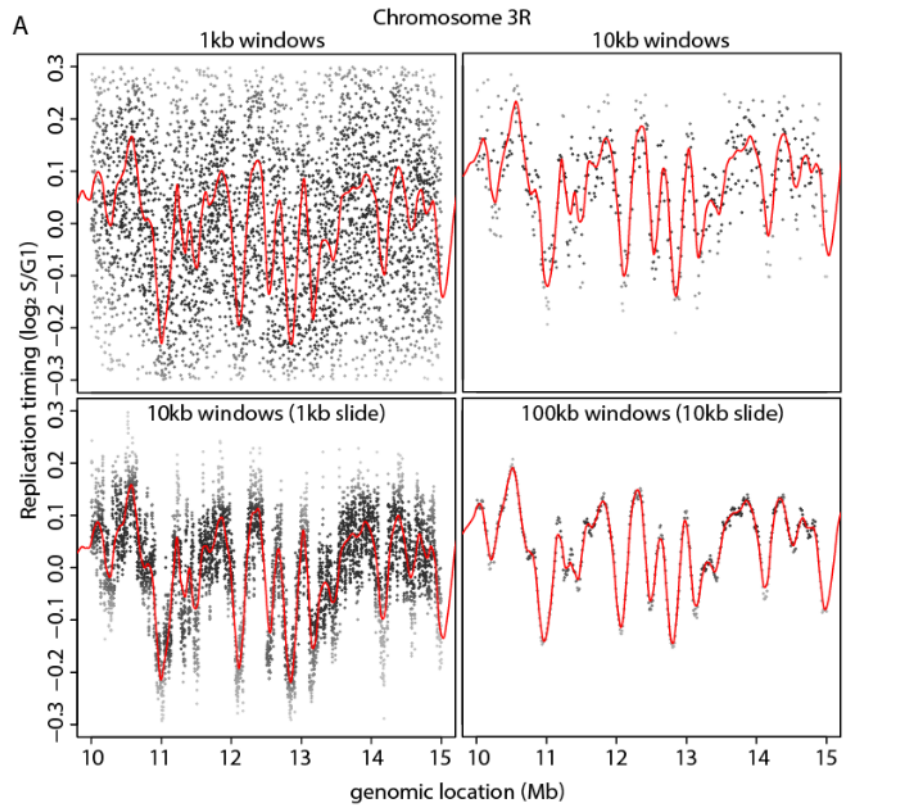

A
1. Sort nuclei
G1 phase gate
S phase gate
G2 phase gate
ungated
Count Density
DNA Content
2. Sequence G1, S, and G2 fractions
Early
Late
3. Replication timing value: log2 (S/G1 ratio)
B
Chromosome 2L
Chromosome 2R
Chromosome 3L
Chromosome 3R
Replication timing log2 (S/G1)
Earlier
Later
genomic location (Mb)
C
Transcript density
Later
Replication timing
Earlier
D
Average ChIP / input
H3K36me3
H3K4me2
H3K4me3
H3K4me1
H3K79me1
H3K27ac
H3K79me2
H3K79me3
H4K20me1
H2Bubi
H3K36me1
H3K23ac
H3K9ac
H3K9me1
H3K9me2
H3K9me3
H3K27me1
H1
H3K27me2
H3K27me3
H3K9acS10P
H4K8ac
H2av
H4K16ac
Early
E/M
Mid
M/L
Late
E
Chromosome 3R
Transposon count
Early
Early/Mid
Mid
Mid/Late
Late
genomic location (Mb)

Figure 2.2. Generation of wild type replication timing profiles.

A) To determine the most appropriate window size for analyzing replication timing, we considered windows of 1kb (top left), 10kb (top right), 10kb with 1kb slide (bottom left), and 100kb with 10kb slide (bottom right). Chr 3R coordinates 10000000-15000000 are shown. A heatscatter of raw $\log_2$ S/G1 values (grey) and LOESS regression line (red) are included. All window sizes yielded highly similar replication timing profiles with 100kb windows and 10kb slide producing raw data that is virtually superimposable with the LOESS regression line. Furthermore, 100kb windows with a 10kb slide fit most closely with the size of replication domains in *Drosophila* (Figure 2.3). Using windows of this large size also reduces complications that might arise from regions of poor mappability when using small windows. B) Scatterplot of the ratio of *H3K9R* G1 reads per million (RPM)/*HWT* G1 RPM on Chr 2 and Chr 3. Differences in replication timing profiles between *H3K9R* and *HWT* are not due to differences in the G1 copy number control. Experiments were performed in collaboration with Taylor Penke.


A
Chromosome 3R
1kb windows
10kb windows
10kb windows (1kb slide)
100kb windows (10kb slide)
Replication timing (log2 S/G1)
genomic location (Mb)

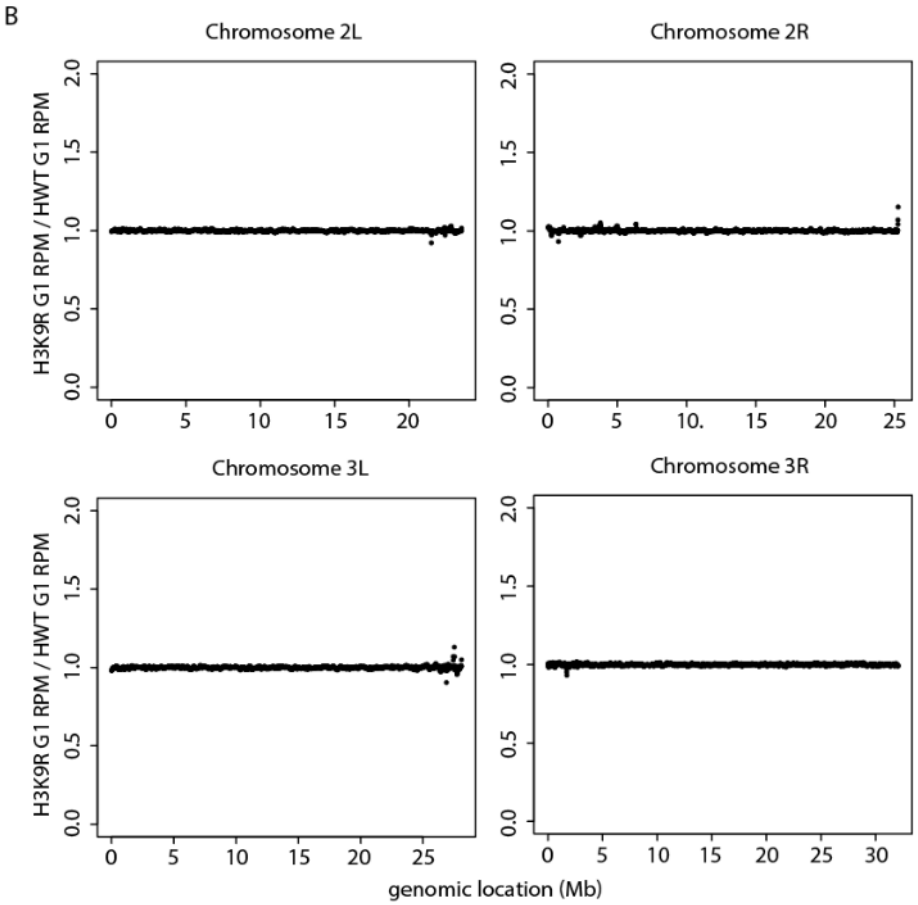

B
Chromosome 2L
Chromosome 2R
Chromosome 3L
Chromosome 3R
H3K9R G1 RPM / HWT G1 RPM
genomic location (Mb)

Figure 2.3. Replication timing in *Drosophila* wing discs correlates with features of active and repressive chromatin.

A) Representative 5Mb region on Chromosome 3R of S/G1 (log_2) replication timing values within 100kb windows with a 10kb slide. RT values are an average of replicate *yw* samples. LOESS regression line is indicated in red. B) LOESS regression line showing average *yw* S/G1 (log_2) replication timing values at 100kb windows using a 10kb slide across Chromosome X and 4 scaffolds. Approximate locations of constitutive heterochromatin (green) and largely euchromatic regions (blue) are indicated (Riddle et al. 2011; Hoskins et al. 2015). C) Histogram of *yw* replication domain sizes. D) Heatscatter plot of *yw* S/G1 (log_2) replication timing values and RNA expression levels within all 10kb windows across the genome with LOESS regression line in black. E) Heatmap of relative H3K9me2 and H3K9me3 enrichment in bins of equally sized RT quintiles (early, early/mid, mid, mid/late, and late) generated using S/G1 (log_2) RT values within 100kb windows and normalized modENCODE H3K9me2/me3 data from third instar larvae. Color indicates average enrichment of all windows within each replication timing quintile. F) Average modENCODE histone PTM enrichment for all 100kb windows within each of the equally sized replication timing quintiles (E=early, E/M=early-mid, M=mid, M/L= mid-late, and L=late). G) Number of transposons within 100kb windows plotted versus genomic location. The color of each dot indicates the replication timing quintile of the window. H) Boxplot of number of transposons within 100kb windows in early (E), early/mid (E/M), mid (M), mid/late (M/L), and late (L) replication timing quintiles. Experiments were performed in collaboration with Taylor Penke.

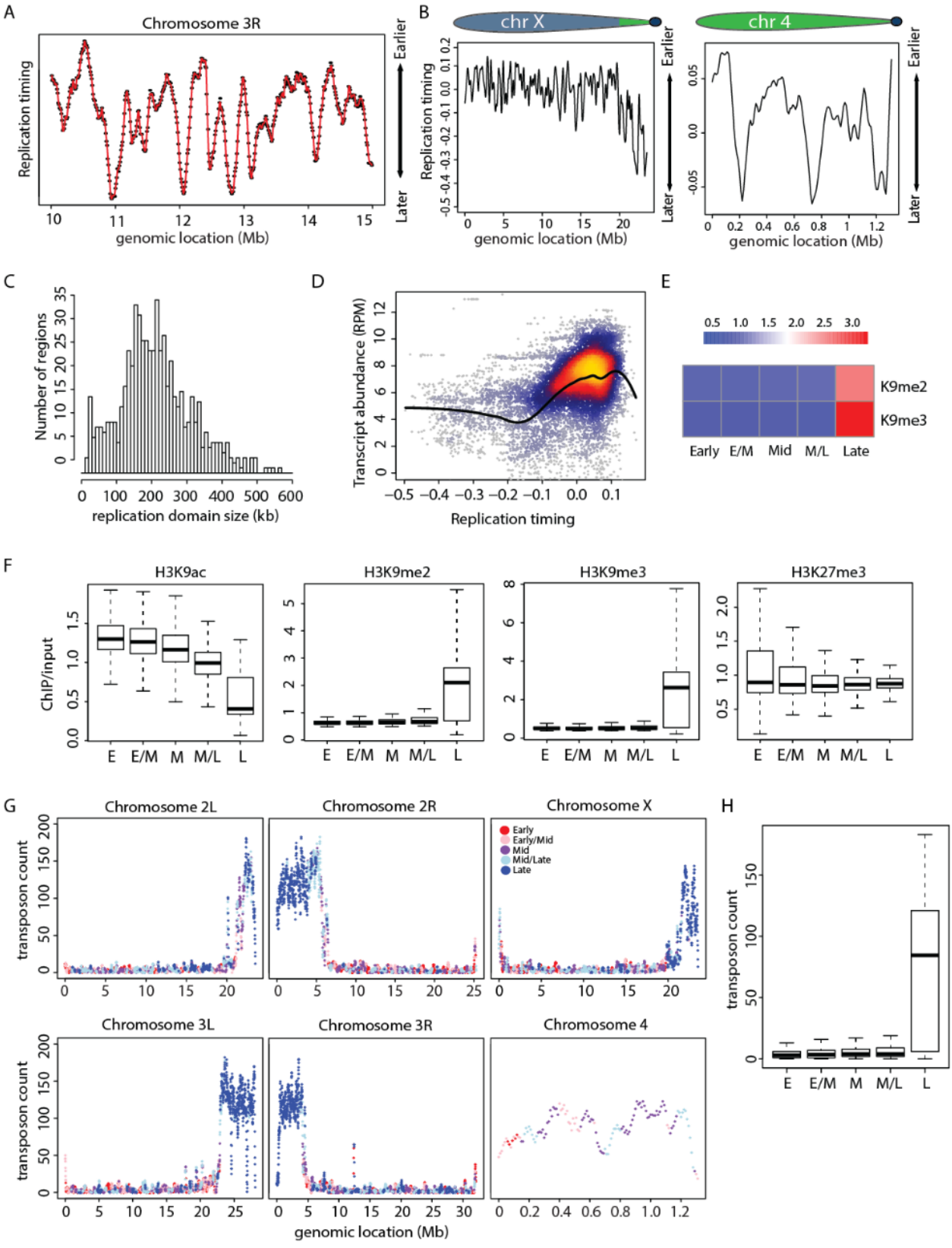
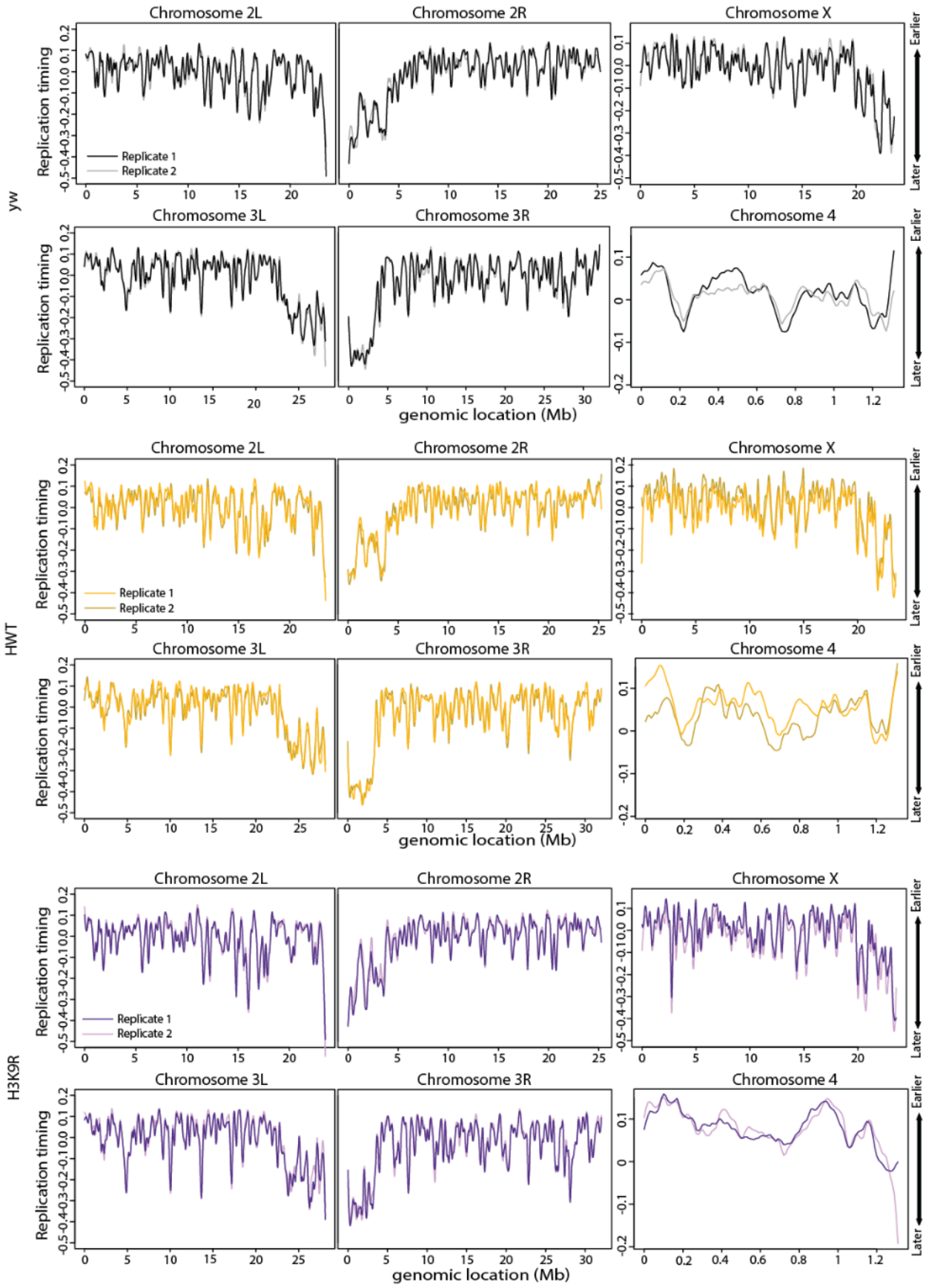

A
Chromosome 3R
Replication timing
Earlier
Later
genomic location (Mb)
B
chr X
chr 4
Replication timing
genomic location (Mb)
C
Number of regions
replication domain size (kb)
D
Transcript abundance (RPM)
Replication timing
E
K9me2
K9me3
Early E/M Mid M/L Late
F
H3K9ac
H3K9me2
H3K9me3
H3K27me3
ChIP/input
E E/M M M/L L
G
Chromosome 2L
Chromosome 2R
Chromosome X
Chromosome 3L
Chromosome 3R
Chromosome 4
Early
Early/Mid
Mid
Mid/Late
Late
transposon count
genomic location (Mb)
H
transposon count

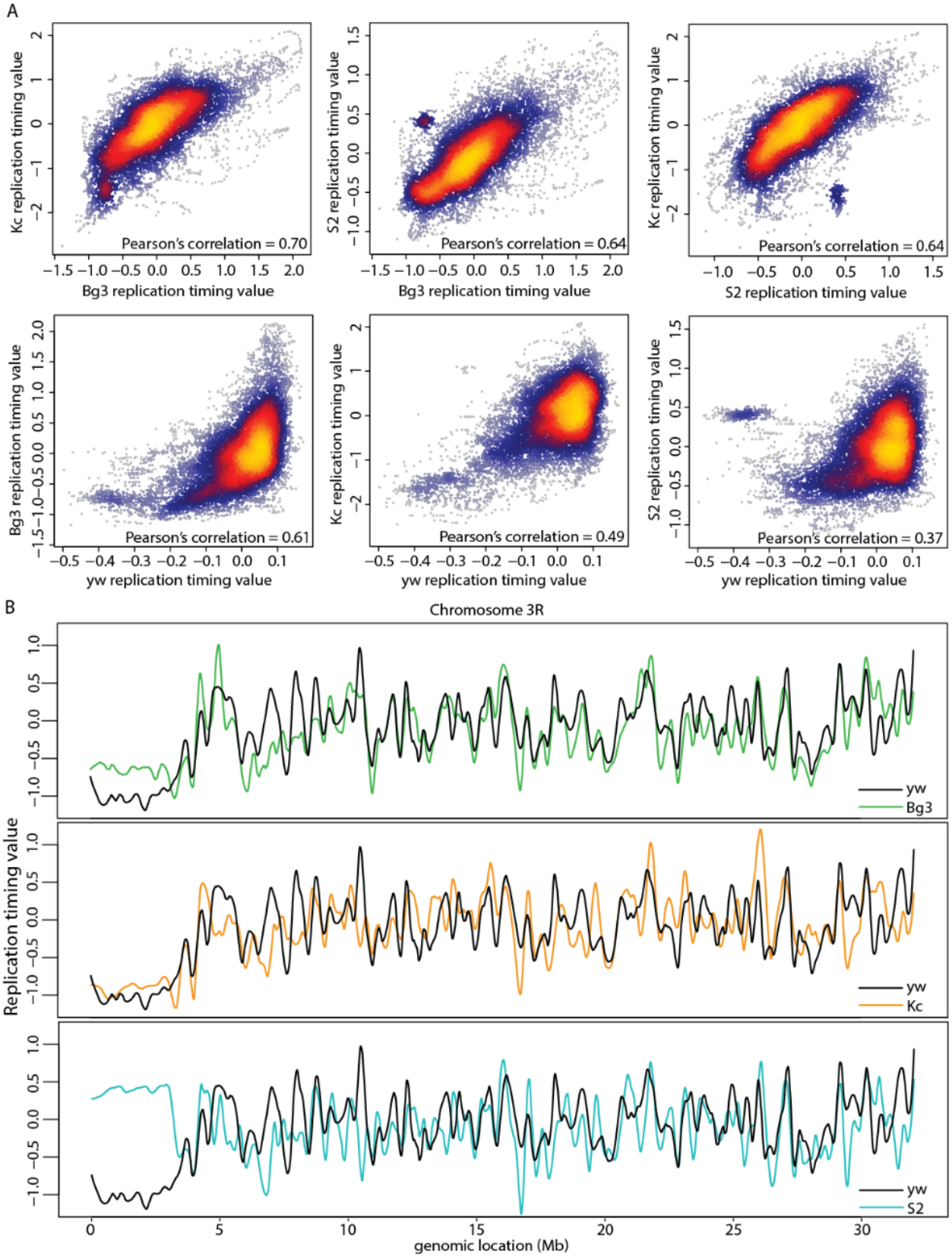
Figure 2.4. Replication timing profiling in *Drosophila* tissue is highly reproducible. Quantile normalized S/G1 (log_2) replication timing values for each replicate for the indicated genotypes were plotted versus genomic coordinate for all major chromosome scaffolds. Each replicate *yw*, *HWT*, and *H3K9R* profile is shown in a different shade of grey, yellow, and purple, respectively. Experiments were performed in collaboration with Taylor Penke.


yw
HWT
H3K9R
Chromosome 2L
Chromosome 2R
Chromosome X
Chromosome 3L
Chromosome 3R
Chromosome 4
Replication timing
Replicate 1
Replicate 2
Earlier
Later
genomic location (Mb)

Our wing disc replication profiles are similar to those previously generated from *Drosophila* cell lines and most closely correlate with RT data obtained from a cell line derived from the same developmental stage as wing discs (Fig 2.5) (Lubelsky et al. 2014). Replication domain sizes ranged from 20kb-570kb (Fig 2.3C), closely matching previous measurements (MacAlpine et al. 2004). Consistent with previous studies in zebrafish embryos and in fly and mammalian cultured cells (Bell et al. 2010; Eaton et al. 2011; Lubelsky et al. 2014; Petryk et al. 2016; Siefert et al. 2017), we found that earlier replication correlates with higher gene density (Fig 2.1C), higher levels of transcription (Fig 2.3D), and the presence of activating histone PTMs such as H3K4me and H3K9ac (Fig 2.1D). In contrast, later replication occurred in gene-poor regions (Fig 2.1C) and was enriched in transposons (Fig 2.1E, Fig 2.3G,H) and repressive histone PTMs, such as H3K9me2/me3 (Fig 2.1D, Fig 2.3E,F). Our RT data revealed that in wild type wing discs the pericentric heterochromatin replicates later than the mostly euchromatic chromosome arms (Fig 2.1B; Fig 2.3B), consistent with prior cytological observations (Taylor 1960; Shermoen et al. 2010). Despite replicating at largely different times on average, both pericentric and euchromatic regions contained earlier and later replicating domains within them, such that the earliest replicating domains in pericentric heterochromatin exhibited similar values to the latest replicating domains on euchromatic chromosome arms. Thus, highly reproducible replication profiles from *Drosophila* tissue can be generated that match general features of replication found in other systems.

Figure 2.5. Wild-type 3rd instar imaginal wing discs and cell culture replication timing profiles are highly correlated.

A) Heatscatter plot of S/G1 (log_2) replication timing value at 100kb windows from *yw* imaginal wing discs and previously generated timing profiles from three *Drosophila* cell culture lines (Kc, S2, and Bg3) (Lubelsky et al. 2014). Top row shows correlations between each of the three cell culture lines, and bottom row shows the correlations between *yw* wing discs and the three cell culture lines. Windows with earlier timing values in S2 cells compared to other cell types are located in the pericentromeres and may be due to copy number differences in these regions. B) Comparison of replication timing profiles between *yw* wing discs and each of the three cell culture lines on Chromosome 3R. Experiments were performed in collaboration with Taylor Penke.


A
Kc replication timing value
Bg3 replication timing value
Pearson's correlation = 0.70
S2 replication timing value
Pearson's correlation = 0.64
Pearson's correlation = 0.64
S2 replication timing value
Bg3 replication timing value
Pearson's correlation = 0.61
yw replication timing value
Kc replication timing value
Pearson's correlation = 0.49
yw replication timing value
S2 replication timing value
Pearson's correlation = 0.37
yw replication timing value
B
Chromosome 3R
Replication timing value
yw
Bg3
yw
Kc
yw
S2
genomic location (Mb)

Replication timing is largely unchanged in *H3K9R* mutants

To determine how chromatin structure influences replication, we first tested if modification of H3K9 determines the difference in RT between heterochromatin and euchromatin. Defining features of heterochromatin are the presence of methylated H3K9 (H3K9me) and Heterochromatin Protein 1a (HP1a). HP1a binds H3K9me and facilitates heterochromatin formation through multimerization of HP1a molecules and recruitment of other factors (Canzio et al. 2011; Larson et al. 2017; Strom et al. 2017). Previously, we showed that *H3K9R* mutants are depleted of H3K9me and HP1a within pericentric heterochromatin (Penke et al. 2016). In addition, we found that loci within the pericentric heterochromatin of *H3K9R* mutants are nucleosome depleted relative to controls, as measured by increased FAIRE-seq signal (Formaldehyde-Assisted Isolation of Regulatory Elements) (Penke et al. 2016).

If increasing chromatin accessibility directly resulted in earlier replication initiation, we would expect large-scale advancement of RT at nucleosome-depleted *H3K9R* pericentromeres. We assigned RT values to 100kb windows tiled 10kb across the genome and used stringent significance thresholds ($p<0.01$ (adjusted for multiple testing), absolute $\log_2$ fold change>0.1; limma) to identify differential RT between *H3K9R* and control. Approximately 97% of the *H3K9R* genome has a similar replication profile compared to control, including much of the pericentric heterochromatin (Fig 2.6A; Fig 2.7). Consistent with these findings, cytological analysis of *H3K9R* imaginal cells revealed colocalization of the 359bp pericentric satellite repeat on the X with late-patterned EdU incorporation at DAPI-bright chromocenters (Fig 2.8A), demonstrating that X pericentric heterochromatin remains late replicating. Late replicating *H3K9R* pericentric heterochromatin is not due to

compensation by H3.3 variant histones because *H3.3K9R*; *H3K9R* mutants incapable of producing any H3K9me also contain late-patterned EdU incorporation at DAPI-bright chromocenters (Fig 2.6C,D). FACS analysis revealed a small but statistically significant decrease in the number of S phase cells in *H3K9R* wing discs, indicating that cell cycle phasing is only slightly perturbed (Fig 2.6E). These data show that H3K9 modification is dispensable for RT across most of the genome and that pericentric heterochromatin lacking HP1a and a closed chromatin configuration generally remains late replicating.

Figure 2.6. Analysis of replication timing in *H3K9R* mutants.

A) Log_2 S/G1 RT values at 100kb windows with 10kb slide for 12x *HWT* (histone wild type; yellow) and 12x *H3K9R* (purple) plotted across Chr 3R. See Fig S5 for other chromosomes. B) ~5 Mb region of the pericentromeric heterochromatin of Chr 3R. Red vertical bars designate significant RT changes between *H3K9R* and *HWT* ($p<0.01$, p value adjusted for multiple testing; absolute log_2 fold change>0.1; limma). C) $H3.3^{WT}$ $H3^{WT}$ and $H3.3^{K9R}$ $H3^{K9R}$ (see supplementary materials for full genotype) first instar brains pulse labeled for 1hr with EdU (yellow) and stained for DNA (blue; DAPI). White arrowheads designate late patterned EdU incorporation. D) Percentage of EdU+ cells with early or late EdU incorporation patterns from ~200 cells per genotype. There is no difference between genotypes ($p>0.05$, Chi-squared test). E) Cell cycle indices for *HWT* (yellow) and *H3K9R* (purple) wing disc cells acquired via FACS (calculated using the Dean-Jett-Fox model). Error bars indicate standard deviation of three experiments (* = $p<0.05$). F) All advanced (red) or delayed (blue) 10kb windows in *H3K9R* mutants were assigned to the nine chromatin states defined in flies (Kharchenko et al. 2011). Shown are the percentage of windows that overlap each chromatin state. G) Average enrichment of modENCODE H3K9ac, me2, me3, and H3K27me3 signal from third instar larvae at 10kb windows of advanced (red), delayed (blue), or randomized set of windows (Celniker et al. 2009). Experiments were performed in collaboration with Taylor Penke.

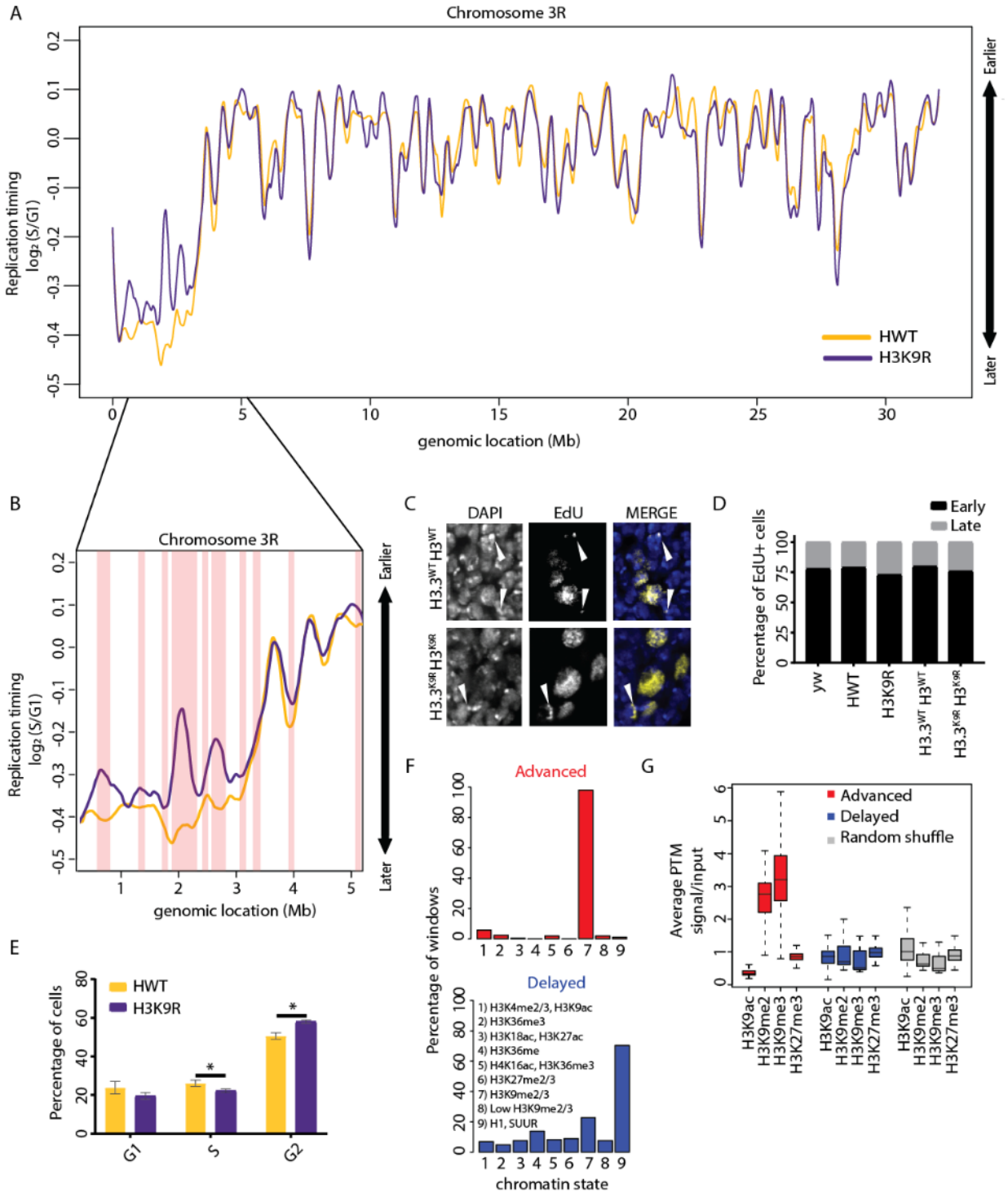

A
Chromosome 3R
Replication timing
log2 (S/G1)
genomic location (Mb)
HWT
H3K9R
Earlier
Later
B
Chromosome 3R
genomic location (Mb)
C
DAPI
EdU
MERGE
H3.3WT H3WT
H3.3K9R H3K9R
D
Early
Late
Percentage of EdU+ cells
yw
HWT
H3K9R
H3.3WT H3WT
H3.3K9R H3K9R
E
HWT
H3K9R
Percentage of cells
G1
S
G2
F
Advanced
Delayed
Percentage of windows
1) H3K4me2/3, H3K9ac
2) H3K36me3
3) H3K18ac, H3K27ac
4) H3K36me
5) H4K16ac, H3K36me3
6) H3K27me2/3
7) H3K9me2/3
8) Low H3K9me2/3
9) H1, SUUR
chromatin state
G
Advanced
Delayed
Random shuffle
Average PTM signal/input
H3K9ac
H3K9me2
H3K9me3
H3K27me3

Figure 2.7. Replication timing profile for *H3K9R* mutants and control.

LOESS regression line applied to S/G1 (log_2) averaged replicates from *HWT* (yellow) and *H3K9R* (purple) plotted across all major chromosome scaffolds at 100kb windows with a 10kb slide. Experiments were performed in collaboration with Taylor Penke.

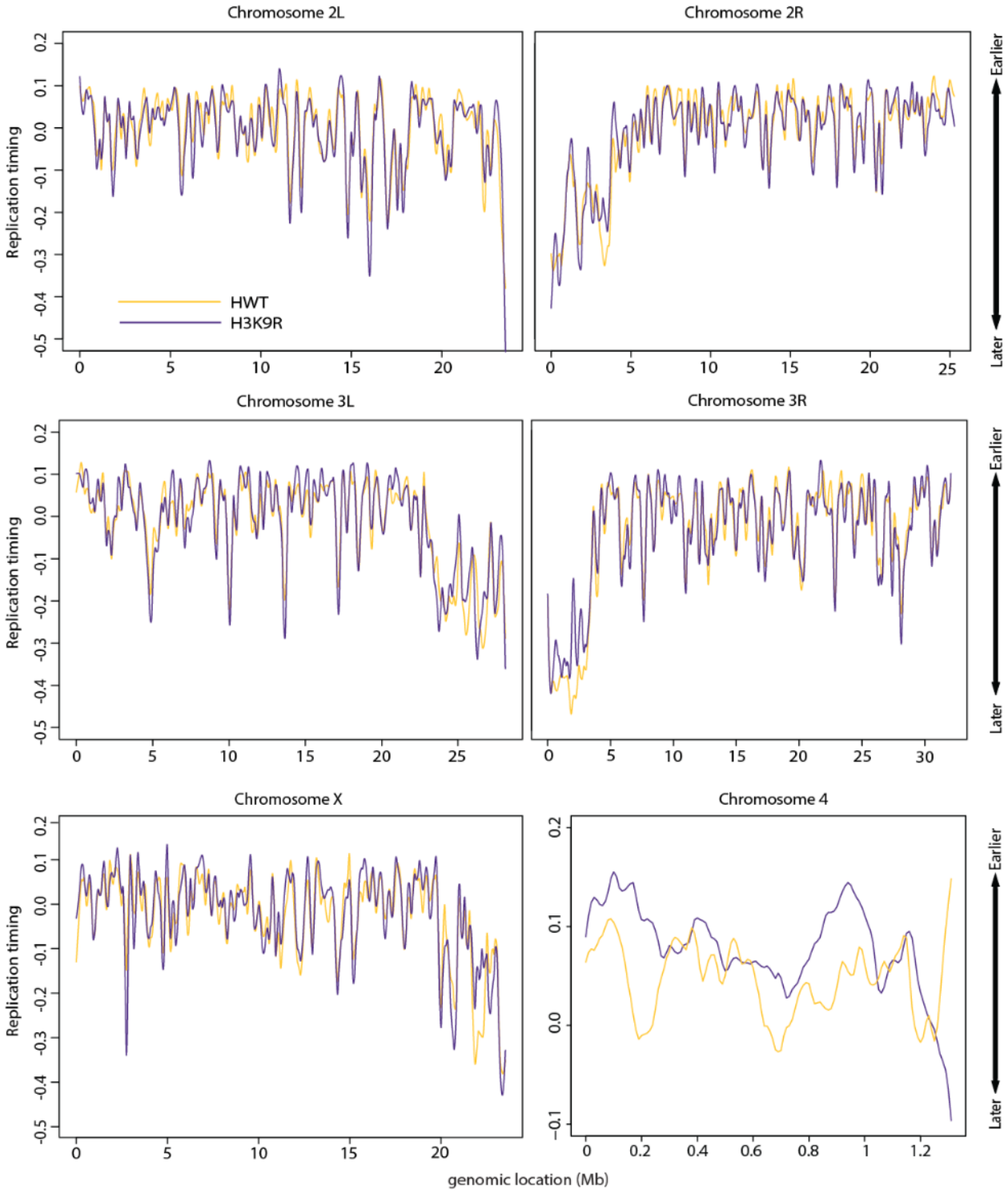
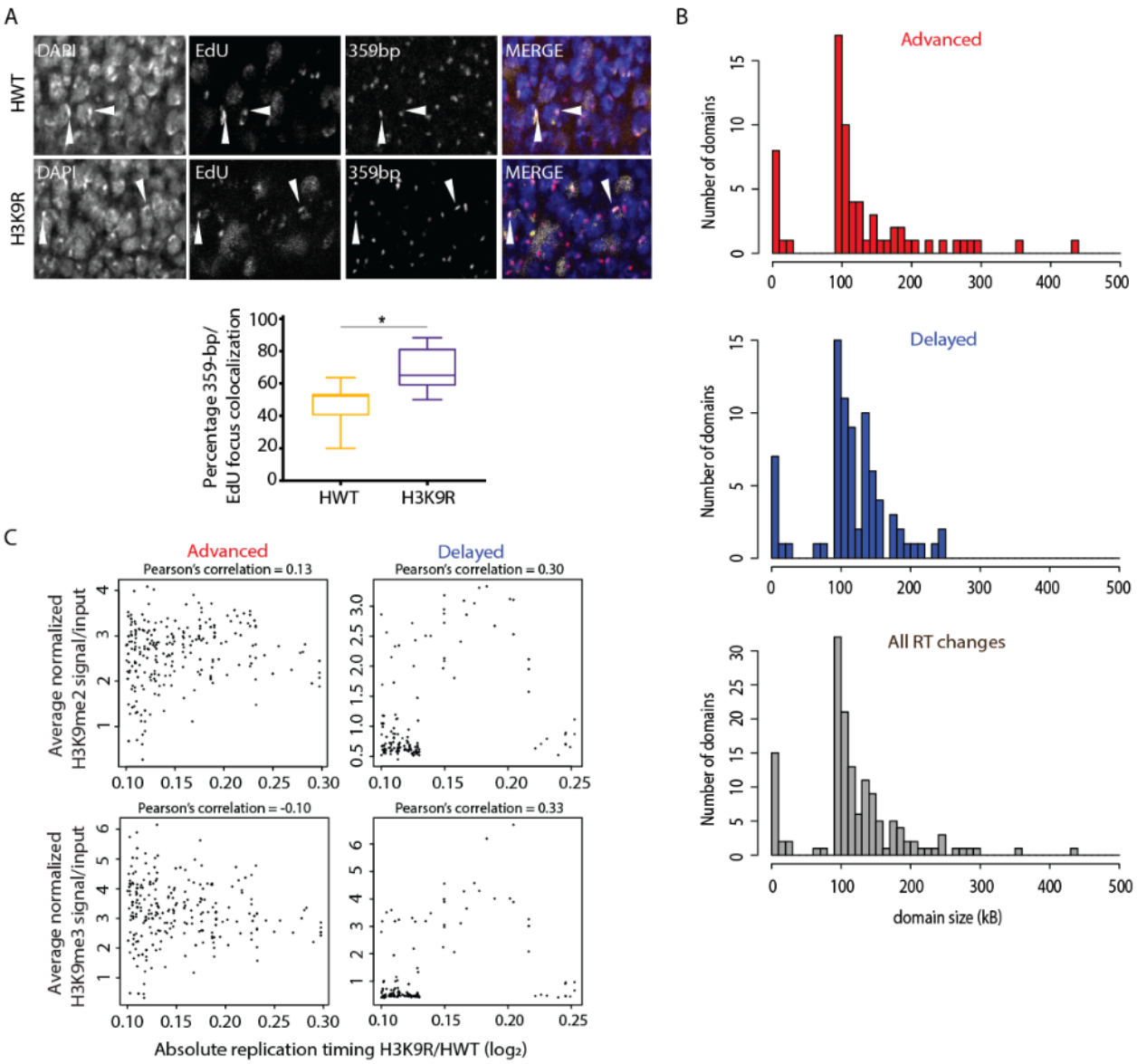

Chromosome 2L
Chromosome 2R
Chromosome 3L
Chromosome 3R
Chromosome X
Chromosome 4
Replication timing
HWT
H3K9R
Earlier
Later
genomic location (Mb)

Figure 2.8. Characterization of altered replication timing in *H3K9R* mutants.

A) *HWT* and *H3K9R* eye imaginal discs labeled for 60' with EdU (yellow), stained for 359-bp FISH probe (X Chromosome pericentromere; magenta), and counterstained with DAPI (blue). White arrow indicates colocalization of EdU late replication focus and 359-bp FISH probe. Shown is a single slice from a Z projection (top). Box plot of the percentage of colocalization between the 359-bp FISH focus and late-patterned EdU focus in *HWT* and *H3K9R* eye imaginal discs (bottom; * = $P < 0.05$, Student's T-test). B) Histogram of the number of domain sizes with advanced (red), delayed (blue), or all replication timing change (grey). C) Correlation analysis of the absolute *H3K9R*/*HWT* $\log_2$ RT fold change versus the average enrichment of H3K9me2 (top) or H3K9me3 (bottom) signal at 10kb windows with significantly advanced (left) or delayed (right) replication. ChIP-seq enrichment was determined from modENCODE datasets from wild-type whole 3rd instar larvae. Experiments were performed in collaboration with Taylor Penke.


A
HWT
H3K9R
DAPI
EdU
359bp
MERGE
Percentage 359-bp/ EdU focus colocalization
HWT
H3K9R
B
Advanced
Delayed
All RT changes
Number of domains
domain size (kB)
C
Advanced
Delayed
Pearson's correlation = 0.13
Pearson's correlation = 0.30
Pearson's correlation = -0.10
Pearson's correlation = 0.33
Average normalized H3K9me2 signal/input
Average normalized H3K9me3 signal/input
Absolute replication timing H3K9R/HWT (log2)

Advanced replication occurs at newly accessible chromatin in *H3K9R* mutants

Despite largely unchanged RT in *H3K9R* mutants, 3% of the genome nevertheless exhibited altered RT (~2% advanced and ~1% delayed). Importantly, these changes do not result from pre-existing copy number differences between the G1 genomes of *H3K9R* and control (Fig 2.2). We used these changes to investigate the relationship between chromatin structure and replication initiation (Fig 2.8B). We found that the majority (82.1%) of earlier replicating 100kb windows in *H3K9R* mutants are located in pericentric heterochromatin (Fig 2.6B) or on the small 4th Chromosome (Fig 2.7), which is primarily heterochromatic (Haynes et al. 2007). Importantly, these changes are unlikely to be caused by changes in the expression of genes encoding replication factors or other protein-coding genes, as the *H3K9R* mutation does not significantly affect their expression (File 2.3; (Penke et al. 2016)). By contrast, 76.2% of later replicating 100kb windows are located along euchromatic chromosome arms (Fig 2.6A; Fig 2.7).

To compare our RT data to other genome features like histone PTMs, we assigned a RT value to non-overlapping 10kb windows across the entire genome (see supplementary materials). Notably, 10kb windows with advanced replication in *H3K9R* mutants are enriched for H3K9me2/me3 in a wild type genome and not for other histone PTMs such as H3K27me3, a marker of facultative heterochromatin (Fig 2.6F,G; Fig 2.8C). This observation suggests that advanced replication is a direct effect of the *H3K9R* mutation, even though most regions enriched for H3K9me2/me3 do not change RT in *H3K9R* mutants. In contrast, delayed replication was not correlated with H3K9me2/me3 and instead occurred preferentially in chromatin environments relatively devoid of histone PTMs, referred to as "Black" chromatin (Fig 2.6F,G; Fig 2.8C) (Filion et al. 2010).

We hypothesized that if chromatin structure directly influences replication, then RT changes should occur at newly accessible chromatin in *H3K9R* mutants. To compare chromatin accessibility and RT in *H3K9R* mutants, we compared FAIRE-seq and RT values at 10kb windows across the genome (Materials and Methods) (Penke et al. 2016). While most pericentric regions included in the current genome assembly are more accessible in *H3K9R* mutants compared to control (Penke et al. 2016), we found that the vast majority (92.9%) of windows with increased FAIRE signal do not display altered RT (Fig 2.9A). Thus, despite established correlations between accessible chromatin and early replication, increasing chromatin accessibility by *H3K9R* mutation does not invariably result in earlier replication.

Importantly, this conclusion does not mean that high chromatin accessibility makes no contribution to early replication. Indeed, nearly all windows (230/243) that exhibit significantly advanced replication in *H3K9R* mutants also have increased FAIRE signal (Fig 2.9A-D; Fig 2.10A,E). This result suggests that a more accessible chromatin environment may be necessary for earlier replication in *H3K9R* mutants. In contrast, most windows with delayed RT exhibit no change in FAIRE signal, suggesting that delayed replication occurs independently from chromatin accessibility changes in *H3K9R* mutants (Fig 2.9A-D; Fig 2.10A,F).

We made similar observations when considering HP1a chromatin binding (Fig 2.10B), which we previously showed is depleted from regions of the *H3K9R* genome that largely overlap regions of increased chromatin accessibility (Fig 2.10C) (Penke et al. 2016). HP1a is depleted at 217 of the 243 10kb windows that advanced RT in *H3K9R* mutants (Fig 2.9C,D; Fig 2.10B,D,F). However, the majority (94.7%) of windows that lose HP1a in

H3K9R mutants do not have altered RT. These results indicate that HP1a loss does not invariably result in advanced replication in *H3K9R* mutants, although it may be necessary. Overall our observations are surprising in that the hallmarks of heterochromatin – high levels of H3K9me and HP1a within a relatively inaccessible chromatin environment – are not necessary for maintaining late replication of most pericentric heterochromatin in animal cells.

Figure 2.9. Open chromatin is permissive to advancement but not delay of replication timing.

A) Heatscatter plot of the *H3K9R*/*HWT* ratio of RT values (log_2 S/G1) versus the *H3K9R*/*HWT* ratio of FAIRE signal at all 10kb windows across the major chromosome scaffolds. 10kb windows with significantly advanced (red) or delayed (blue) RT are indicated. Darker color indicates higher density of windows. B) Cumulative count of advanced (red) or delayed (blue) 10kb windows ordered by increasing FAIRE signal in *H3K9R* compared to *HWT*. C) Heatscatter plot of the *H3K9R*/*HWT* ratio of HP1a ChIP signal versus the *H3K9R*/*HWT* ratio of FAIRE signal at all 10kb windows across the major chromosome scaffolds. D) Venn-diagram of all 10kb windows with significantly altered FAIRE or HP1a signal in *H3K9R* compared to *HWT* ($p<0.01$; edgeR). For all panels, significantly different RT was determined as $p<0.05$, log_2 fold change>0.1 using limma. Experiments were performed in collaboration with Taylor Penke.

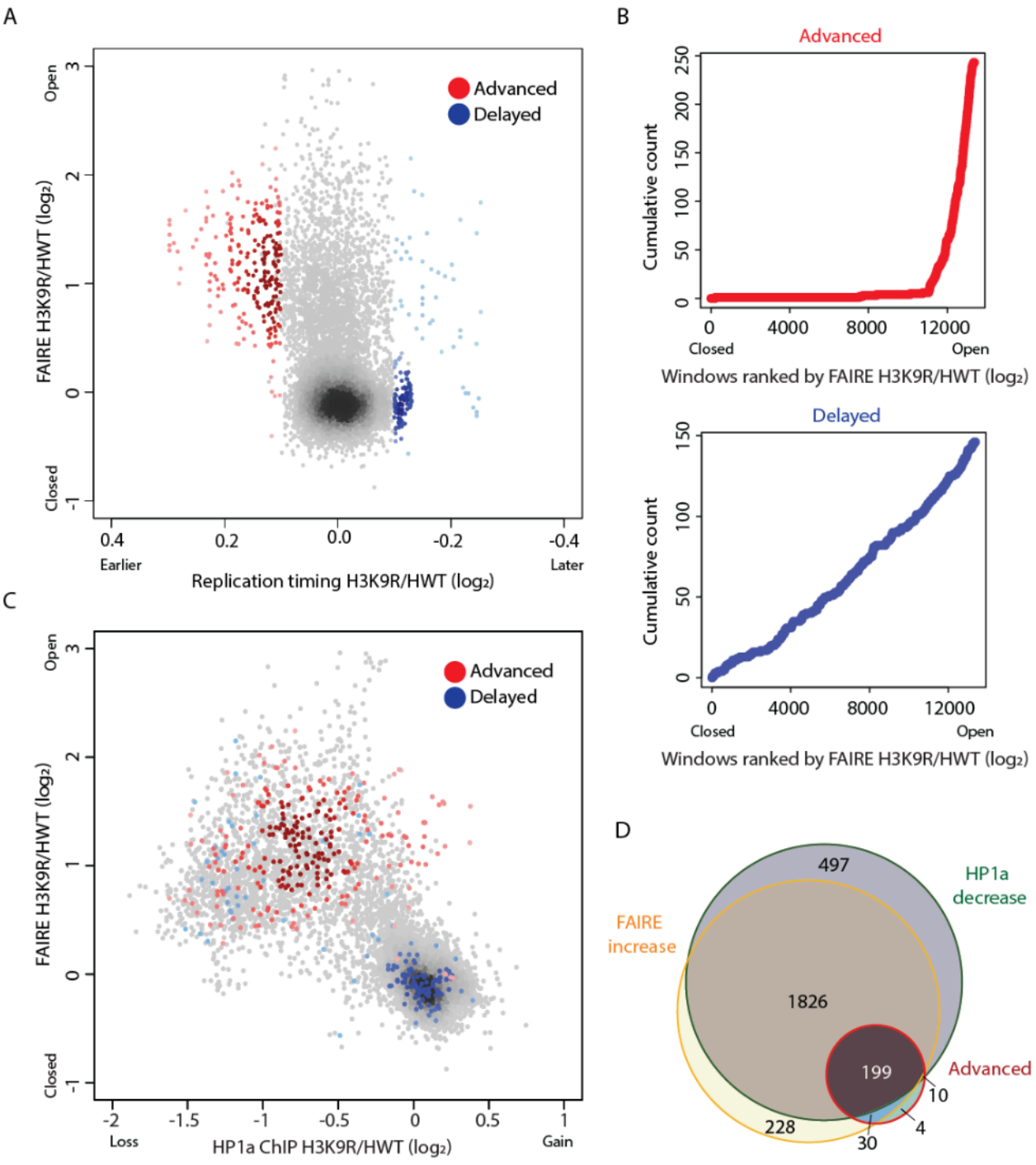
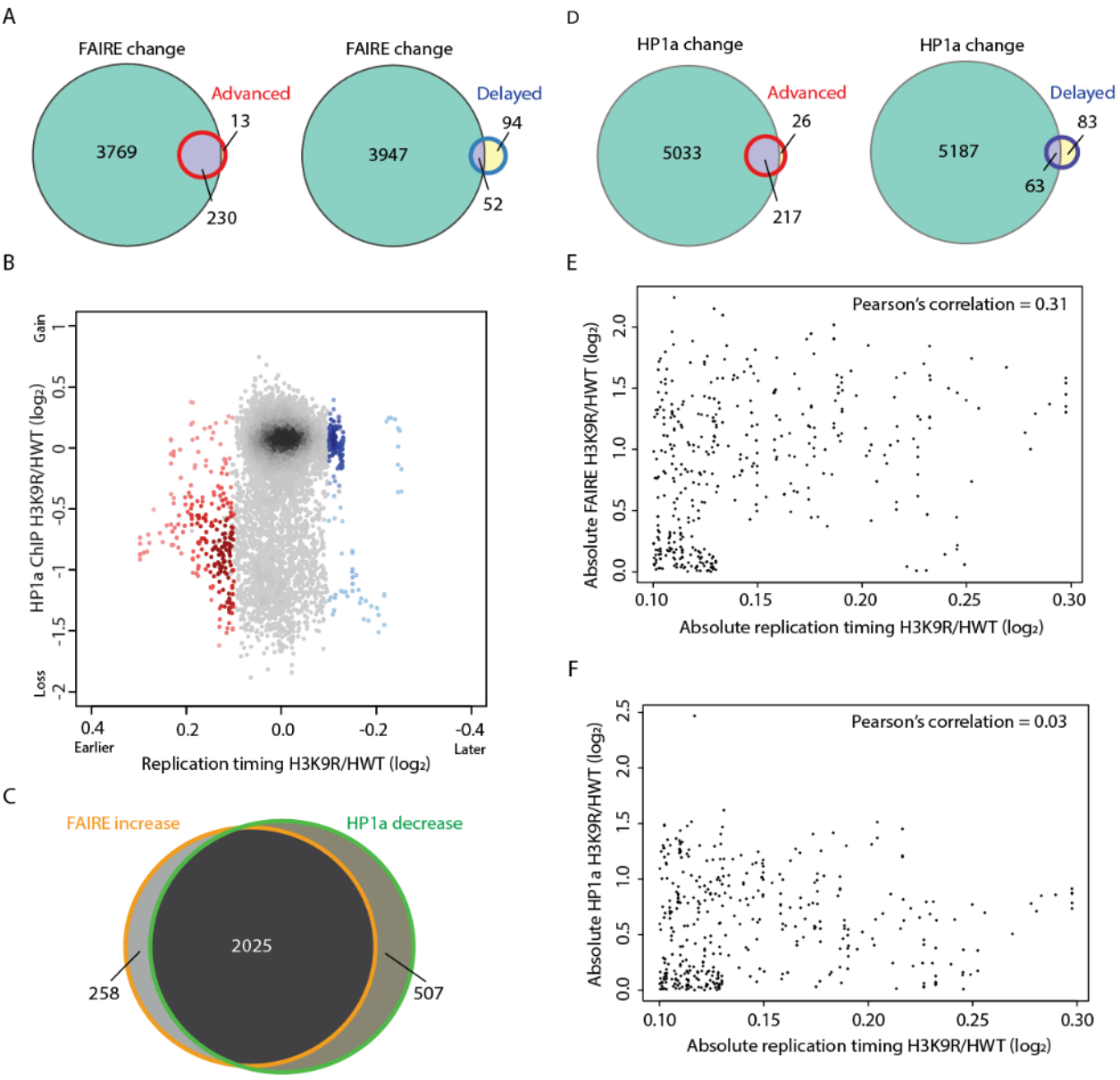

A
Advanced
Delayed
FAIRE H3K9R/HWT (log2)
Open
Closed
3
2
1
0
-1
0.4
0.2
0.0
-0.2
-0.4
Earlier
Later
Replication timing H3K9R/HWT (log2)
B
Advanced
Cumulative count
250
200
150
100
50
0
0
4000
8000
12000
Closed
Open
Windows ranked by FAIRE H3K9R/HWT (log2)
Delayed
Cumulative count
150
100
50
0
0
4000
8000
12000
Closed
Open
Windows ranked by FAIRE H3K9R/HWT (log2)
C
Advanced
Delayed
FAIRE H3K9R/HWT (log2)
Open
Closed
3
2
1
0
-1
-2
-1.5
-1
-0.5
0
0.5
1
Loss
Gain
HP1a ChIP H3K9R/HWT (log2)
D
497
HP1a
decrease
FAIRE
increase
1826
199
Advanced
10
228
4
30

Figure 2.10. Disrupting heterochromatin does not always result in altered replication. A) Venn-diagram of 10kb windows with significantly altered FAIRE signal and significantly advanced or delayed replication in *H3K9R* mutants compared to control. B) Heatscatter plot of the *H3K9R*/*HWT* ratio of normalized replication timing values (S/G1 (log_2)) plotted versus the *H3K9R*/*HWT* ratio of normalized HP1a ChIP signal at all 10kb windows across the major chromosome scaffolds. C) Venn-diagram of 10kb windows with significantly increased FAIRE signal and decreased HP1a ChIP signal. D) Venn-diagram of 10kb windows with significantly altered HP1a ChIP signal and significantly advanced or delayed RT. E-F) Absolute change in FAIRE signal (E) or HP1a ChIP signal (F) between *H3K9R* mutants and controls plotted versus the absolute change in replication timing between the two genotypes. The magnitude of altered chromatin accessibility or HP1a localization is not correlated with the magnitude of replication timing change. Experiments were performed in collaboration with Taylor Penke.


A
FAIRE change
3769
Advanced
13
230
FAIRE change
3947
Delayed
94
52
D
HP1a change
5033
Advanced
26
217
HP1a change
5187
Delayed
83
63
B
HP1a ChIP H3K9R/HWT (log2)
Gain
Loss
Earlier
Later
Replication timing H3K9R/HWT (log2)
E
Pearson's correlation = 0.31
Absolute FAIRE H3K9R/HWT (log2)
Absolute replication timing H3K9R/HWT (log2)
C
FAIRE increase
HP1a decrease
2025
258
507
F
Pearson's correlation = 0.03
Absolute HP1a H3K9R/HWT (log2)
Absolute replication timing H3K9R/HWT (log2)

Elevated transposon expression accompanies advanced replication in *H3K9R* mutants

We next considered the transcriptional activity of domains of altered replication. We compared our newly generated RT profiles with our previously generated wing disc transcriptome profiles from *H3K9R* and control (Penke et al. 2016). We focused on transcripts (genes or transposons) most likely to drive RT changes by identifying the transcript that was most significantly different within each 10kb window between *H3K9R* and control (i.e. the transcript with the lowest p-value in differential expression analysis; edgeR). We then compared the fold change of this transcript to the RT value assigned to the same 10kb window. We found that only a small fraction (6.8%) of the 3,371 10kb windows containing a transcript with a significant expression change also exhibited a RT change (Fig 2.11A,B; Fig 2.12A). This observation indicates that, despite strong correlations between active transcription and early replication (MacAlpine et al. 2004; Liu et al. 2012; Lubelsky et al. 2014; Rivera-Mulia and Gilbert 2016), transcriptional activity and RT are separable. Conversely, we found that the majority (76.5 %) of windows with advanced replication in *H3K9R* mutants exhibited a change in gene expression (Fig 2.11A,B; Fig 2.12B). Because most (97.3%) changes were increases in expression, we speculate that transcription might promote early replication initiation in pericentric heterochromatin. Similar results were obtained by using the average expression change of all transcripts that overlap each window with advanced RT, rather than the transcript with the most significant change in expression across the window (Fig 2.12G).

Windows with advanced replication in *H3K9R* mutants have a high transposon density, unlike delayed windows which are gene-rich (Fig 2.11D; Fig 2.13A). Low sequence mappability of most transposons likely inhibited our ability to detect all transcriptional

changes within advanced replication domains (Fig 2.12C). Therefore, we also identified transposons belonging to families that were differentially expressed between *H3K9R* and control (Materials and Methods; Fig 2.11C; Fig 2.12C-F). All 243 windows of advanced replication in *H3K9R* mutants contain either a transposon belonging to a family that was differentially expressed in *H3K9R* compared to control (96.4%) or that neighbored a window containing multiple differentially expressed transcripts (Fig 2.11C,D). Although we cannot determine whether individual transposons within all 243 advanced windows changed expression, these data suggest that altered transcription may promote advancement of replication in *H3K9R* mutants.

Along with transposon enrichment (Fig 2.13A,B), advanced replication domains in *H3K9R* mutants are normally enriched for H3K9me2/me3 (Fig 2.13C) and exhibited a lower GC content (Fig 2.13D) compared to domains of increased chromatin accessibility or increased RNA expression with unaltered replication (FAIRE only and RNA only, respectively). Although transposon density distinguished advanced domains, the majority of domains with altered transposon expression have no change in RT (Fig 2.11C; Fig 2.12D). Therefore, we surmise that altered transposon expression is necessary, but additional events must occur within accessible chromatin to advance replication.

Our data thus far indicate that increased chromatin accessibility and gene expression act upstream of advanced replication within pericentric heterochromatin in *H3K9R* mutants (Fig 2.11D). To further understand the relationship between transcription and DNA replication within transcriptionally active euchromatin, we investigated the *Drosophila* male X Chromosome, which replicates earlier in males than in females (Schwaiger et al. 2009; Bell et al. 2010; Lubelsky et al. 2014). In addition, a two-fold hyper-activation of gene

expression from the male X results in matched X-linked gene expression between XY males and XX females (Kuroda et al. 2016). We therefore generated a replication dependent (RD) histone genotype (*H4K16R*) predicted to disrupt dosage compensation and determined the effect on gene expression and RT and whether these two processes could be uncoupled.

Figure 2.11. Altered transposon expression occurs at advanced replication domains in *H3K9R* mutants.

A) Heatscatter plot of the *H3K9R*/*HWT* ratio of RT values (log_2 S/G1) plotted versus the *H3K9R*/*HWT* ratio of RNA-seq signal at all 10kb windows across major chromosome scaffolds. RNA-seq differences were determined based on the transcript with the lowest p-value across the 10kb window. 10kb windows with significantly advanced (red) and delayed (blue) RT are indicated ($p<0.05$, log_2 fold change>0.1; limma). B) Histogram of the number of differentially expressed transcripts in 10kb windows of advanced replication (red; left). Venn-diagram comparing the number of windows with differentially expressed transcripts and number of windows with advanced replication (right). C) Histogram of the number of transposons belonging to a differentially expressed transposon family in 10kb windows of advanced replication (red; left). Venn-diagram comparing the number of windows with a transposon belonging to a differentially expressed transposon family to the number of windows with advanced replication (right). D) Browser shot of a 10kb window (Chr 3R-2130000-2140000) with advanced replication. *HWT* (yellow) and *H3K9R* (purple) FAIRE-seq, HP1a ChIP-seq, and RNA-seq data plotted in the context of mappability, genes, and transposons. Experiments were performed in collaboration with Taylor Penke.

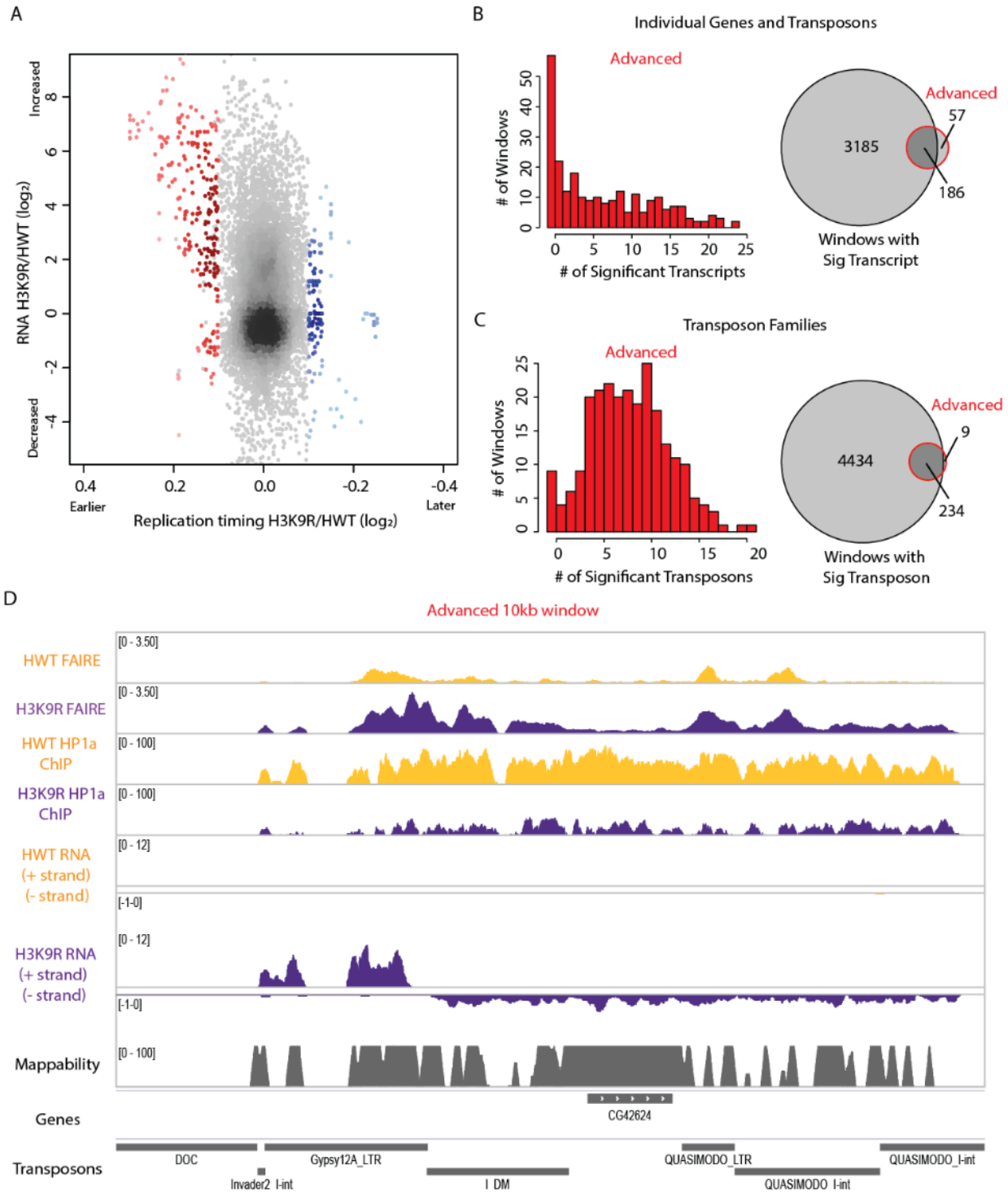

A
Increased
Decreased
RNA H3K9R/HWT (log2)
8
6
4
2
0
-2
-4
0.4
0.2
0.0
-0.2
-0.4
Earlier
Later
Replication timing H3K9R/HWT (log2)
B
Individual Genes and Transposons
Advanced
# of Windows
0
10
20
30
40
50
0
5
10
15
20
25
# of Significant Transcripts
Advanced
57
3185
186
Windows with
Sig Transcript
C
Transposon Families
Advanced
# of Windows
0
5
10
15
20
25
0
5
10
15
20
# of Significant Transposons
Advanced
9
4434
234
Windows with
Sig Transposon
D
Advanced 10kb window
HWT FAIRE
[0 - 3.50]
H3K9R FAIRE
[0 - 3.50]
HWT HP1a
ChIP
[0 - 100]
H3K9R HP1a
ChIP
[0 - 100]
HWT RNA
(+ strand)
(- strand)
[0 - 12]
[-1-0]
H3K9R RNA
(+ strand)
(- strand)
[0 - 12]
[-1-0]
Mappability
[0 - 100]
Genes
CG42624
Transposons
DOC
Gypsy12A_LTR
QUASIMODO_LTR
QUASIMODO_I-int
Invader2_I-int
I_DM
QUASIMODO_I-int

Figure 2.12. Regions of advanced replication in *H3K9R* mutants exhibit altered transposon expression.

A) Histogram of the number of differentially expressed transcripts in 10kb windows of delayed replication (blue; left). Venn-diagram comparing the number of windows with differentially expressed transcripts and number of windows with delayed replication (right). B) Venn-diagram comparing the number of advanced windows in *H3K9R* mutants compared to control containing a differentially expressed transposon and/or a differentially expressed transcript. C) Genome browser shot of a 10kb window with significantly advanced replication in *H3K9R* mutants but no detectable accompanying change in RNA expression via edgeR analysis. FAIRE-seq, HP1a ChIP-seq, or RNA-seq signal are shown for *H3K9R* (purple) and *HWT* (yellow) samples. Note the low mappability of this region due to high transposon density. Red transposons indicate individual transposons belonging to a family that is differentially expressed in *H3K9R* mutants. Browser shot provides a representative example of transcriptional changes that are likely occurring but cannot be directly examined due to low mappability. D) Venn-diagram comparing the number of windows with a differentially expressed transposon to the number of windows with advanced replication (see also Figure 4C). Because high transposon density (Figure 2.12A,B) and low sequence mappability of these regions likely masked our ability to detect transcriptional changes, we examined expression levels of transposon families rather than individual transposons (Materials and Methods). Counts from individual transposons were summed based on RepeatMasker categorization of transposon families. E) MA plot showing differential expression of transposon families between *HWT* control and *H3K9R* samples. Each dot represents a transposon family with red indicating statistical significance as determined by edgeR ($p<0.01$; see Materials and Methods). Blue lines indicate two-fold change. F)

Histograms in the top left panel show the number of transposons belonging to a family that is differentially expressed in *H3K9R* mutants compared to control at 57 10kb windows that exhibited advanced replication in *H3K9R* mutants but no initially detected transcriptional change (see also Venn diagram in Figure 4B). Bottom left panel shows number of transposons bellowing to a differentially expressed family at all 10kb windows that exhibit a transcriptional change but no replication timing change (RNA only). Histograms in right panel show the number of transposons at the 57 advanced windows (top) and RNA only windows (bottom). 52 of the 57 windows we did not initially score as having changed expression (Figure 4B) contained at least one transposon belonging to a family with significantly altered expression in *H3K9R* mutants. The remaining 5 windows were surrounded by 10kb windows containing several transposons with significantly altered expression. These data suggest that altered transcription is necessary for advanced replication in *H3K9R* mutants. G) Scatterplot of RNA RPM at significantly advanced (top, red), delayed (middle, blue), or a randomized set of 10kb windows (bottom, black) in *H3K9R* versus control. RNA expression increases in windows that advance RT in *H3K9R* mutants compared to control. Experiments were performed in collaboration with Taylor Penke.

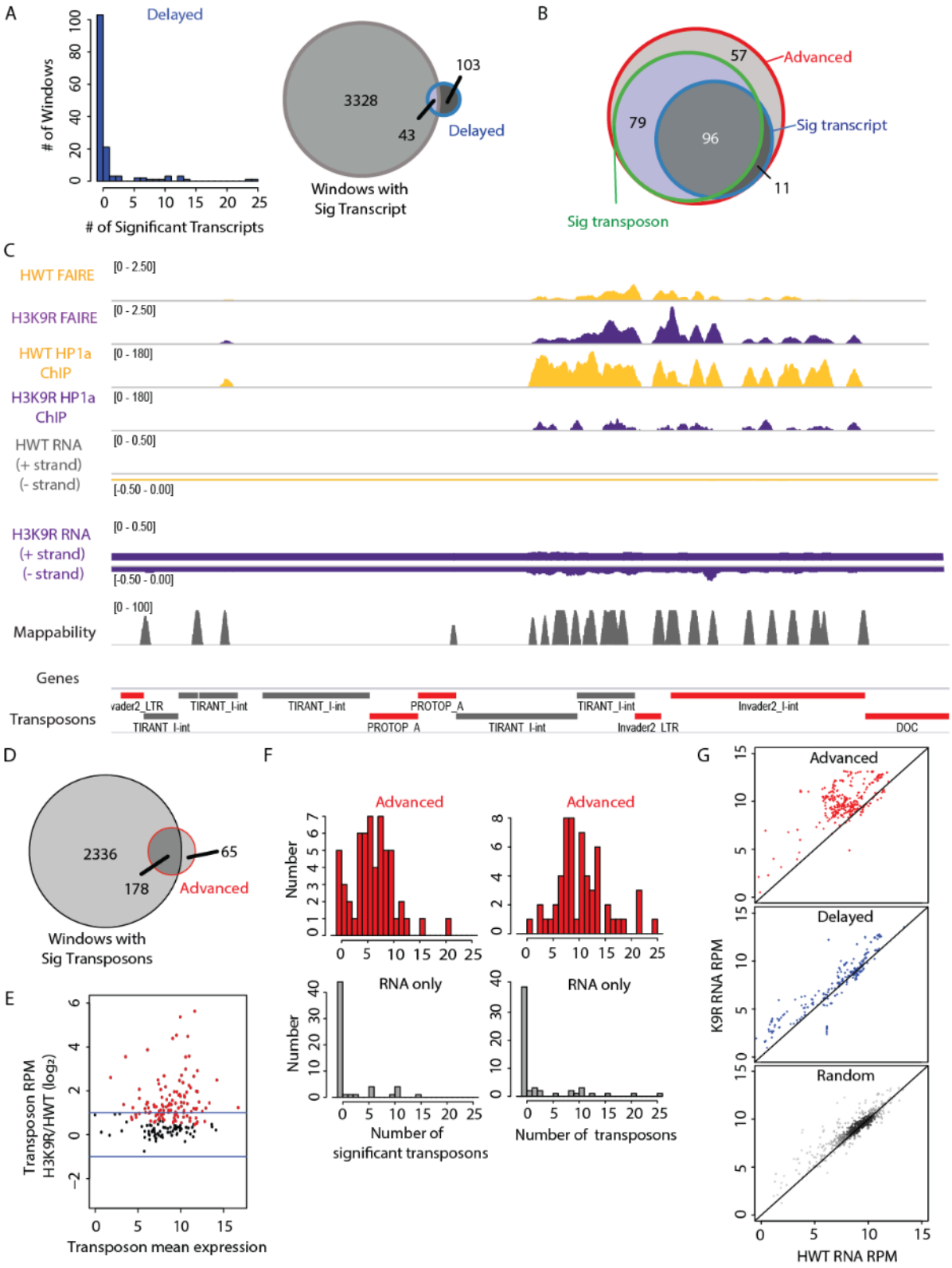
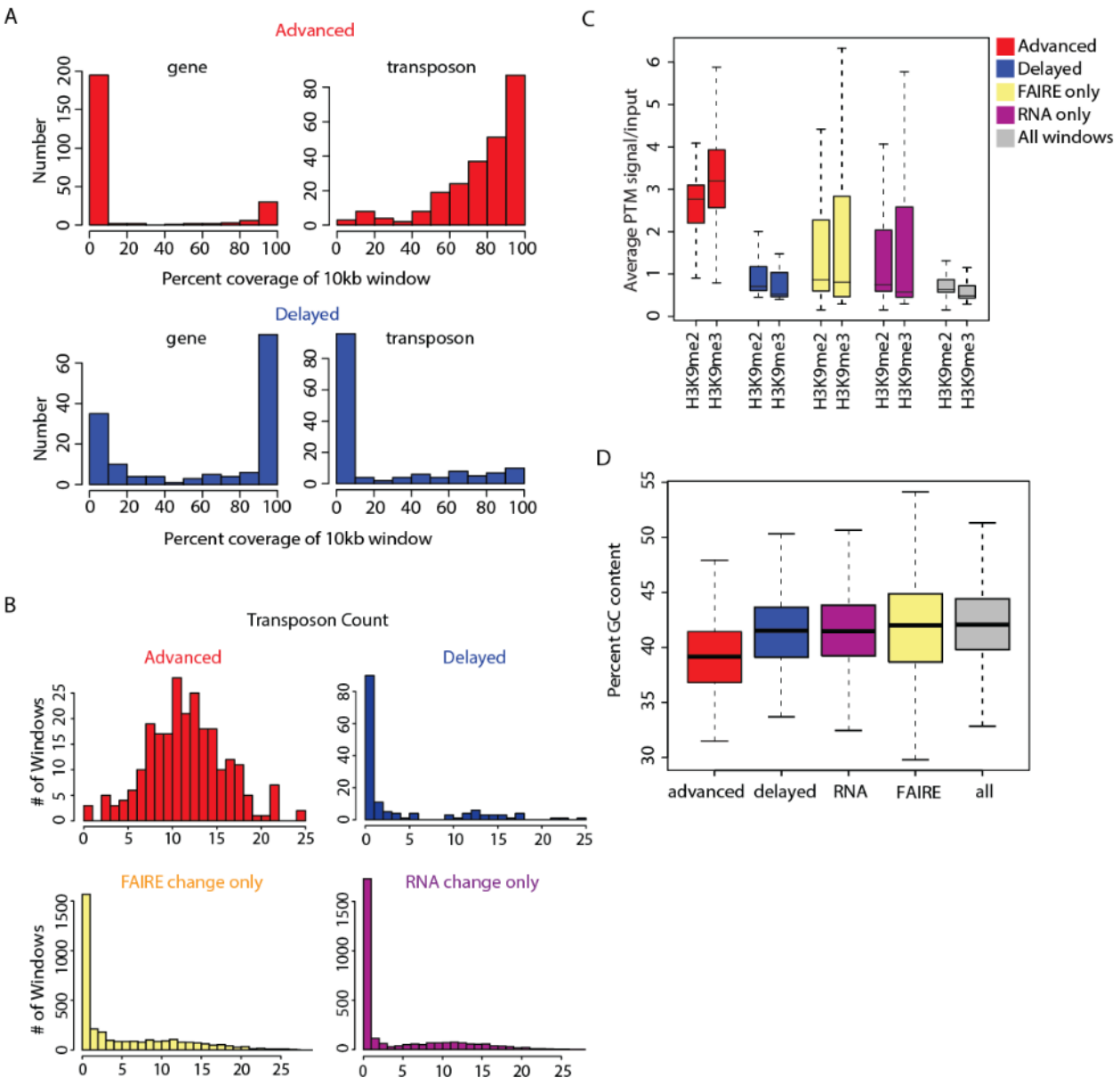

A
Delayed
# of Windows
# of Significant Transcripts
3328
103
43
Delayed
Windows with
Sig Transcript
B
57
Advanced
79
96
Sig transcript
11
Sig transposon
C
HWT FAIRE
H3K9R FAIRE
HWT HP1a ChIP
H3K9R HP1a ChIP
HWT RNA (+ strand) (- strand)
H3K9R RNA (+ strand) (- strand)
Mappability
Genes
Transposons
TIRANT_I-int
PROTOP_A
Invader2_I-int
Invader2_LTR
DOC
D
2336
65
178
Advanced
Windows with
Sig Transposons
E
Transposon RPM H3K9R/HWT (log2)
Transposon mean expression
F
Advanced
RNA only
Number
Number of
significant transposons
Number of transposons
G
Advanced
Delayed
Random
K9R RNA RPM
HWT RNA RPM

Figure 2.13. Transposon density and H3K9me2/me3 status are distinguishing features of regions with advanced replication.

A) Histogram of the number of significantly advanced (red) or delayed (blue) 10kb windows within 10 bins representing two categories: the percentage of each window covered by genes (left panels) or transposons (right panels). B) Histogram of the number of transposons in 10kb windows of advanced replication (red), delayed replication (blue), FAIRE change without replication change (FAIRE only), RNA change without replication change (RNA only), and all 10kb windows. C) Average enrichment of modENCODE H3K9me2 and H3K9me3 signal from wild-type whole third instar larvae at 10kb windows within the categories described in panel B. D) Boxplot of the percent GC content of 10kb windows within the categories described in panel B. Experiments were performed in collaboration with Taylor Penke.


A
Advanced
gene
transposon
Number
Percent coverage of 10kb window
Delayed
gene
transposon
Number
Percent coverage of 10kb window
B
Transposon Count
Advanced
Delayed
# of Windows
FAIRE change only
RNA change only
# of Windows
C
Average PTM signal/input
Advanced
Delayed
FAIRE only
RNA only
All windows
H3K9me2
H3K9me3
D
Percent GC content
advanced
delayed
RNA
FAIRE
all

H4K16 is necessary for hyper-expression of the *Drosophila* male X Chromosome

The *Drosophila* dosage compensation mechanism is mediated by the Male-Specific Lethal (MSL) complex, which specifically localizes to and promotes higher gene expression from the male X. The MSL complex includes MOF, a histone acetyltransferase that acetylates lysine 16 of histone H4, resulting in higher levels of H4K16ac on male X Chromosomes relative to autosomes or the female X (Hilfiker et al. 1997; Smith et al. 2000; Gelbart et al. 2009). Furthermore, hyper-acetylation of H4K16 correlates with increased chromatin accessibility of the male X (Bell et al. 2010). These data suggest that H4K16ac is required for dosage compensation in flies. In accordance with these findings, MOF mutations cause a male-specific lethal phenotype; however, MOF performs both H4K16-dependent and -independent functions (Hilfiker et al. 1997; Buscaino et al. 2003; Sykes et al. 2006). A requirement for H4K16 in dosage compensation, therefore, has not been directly tested.

Similar to observations made using mutations in MOF and other MSL complex members (Lucchesi 1998), we found that male viability is significantly reduced in zygotic, RD *H4K16R* mutants (Fig 2.14A). However, unlike mutation of MSL complex members, which causes fully penetrant male lethality (Lucchesi 1998), these *H4K16R* males can develop to adulthood (Fig 2.14A), although they eclose later than their female siblings. A further reduction in male viability occurred when both maternal and zygotic sources of histones were RD *H4K16R* mutant (Fig 2.14A). The *Drosophila* genome also contains a single copy replication-independent *His4r* gene, which is not located in the RD gene cluster but encodes an identical H4 protein. Combining the RD *H4K16R* zygotic genotype with a CRISPR-derived homozygous deletion of *His4r* resulted in complete male lethality (Fig 2.14A). We therefore conclude that H4K16 function is required for male development. In

contrast, females of all these *H4K16R* genotypes are viable, indicating that H4K16 modification is not generally required for organismal viability.

We next performed gender-specific total RNA-seq from replication dependent *H4K16R* and control wing discs, generated transcriptomes (Cufflinks), and identified differentially expressed transcripts between *H4K16R* males and females and their respective controls (Trapnell et al., 2012). We observed 1789 differentially expressed transcripts (608 increased and 1181 decreased) in *H4K16R* males relative to control males and 105 differentially expressed transcripts in *H4K16R* females relative to control females (39 increased and 66 decreased) indicating that the *H4K16R* effect on gene expression is greater in males than in females ($p<0.05$, edgeR; Fig 2.14B). Of the 1181 genes with decreased expression in *H4K16R* males, 72% are located on the X. In addition, the majority (92%) of the down-regulated, X-linked genes in *H4K16R* males have a $\log_2$ fold change less than 1, which would be expected for a disruption in X Chromosome dosage compensation. In contrast, only 3.6% of genes with increased expression in *H4K16R* males are on the X.

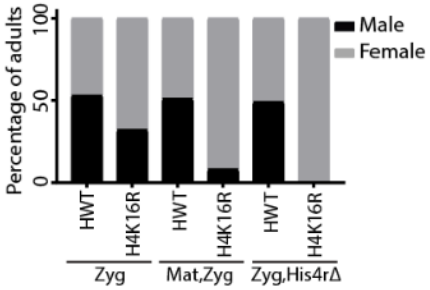
We further examined chromosome-specific differential gene expression by assessing transcript abundance for the X separately from autosomes. We compared our *H4K16R* wing disc RNA-seq data to previously published data from the male *Drosophila* S2 cell line in which MSL2 or MOF had been depleted by RNAi (Zhang et al. 2010). Similar to MSL2 or MOF knockdown, global transcript abundance is decreased for genes on the X in *H4K16R* male wing disc cells compared to control, but not for genes located on the autosomes or the female X ($p<0.05$; Fig 2.14C). Importantly, 10kb windows containing a significantly decreased transcript from the *H4K16R* male X are enriched in a wild type genome for H4K16ac ($p<0.05$; Fig 2.14D). By contrast, windows in *H4K16R* males from autosomes

containing a decreased transcript, those with an increased transcript (either from the X or the autosomes), or all windows with a transcriptional change were not normally enriched for H4K16ac (Fig 2.14D). These data directly demonstrate that H4K16 is a critical component of the *Drosophila* dosage compensation machinery. Moreover, this residue is not required for basal genome function, as female gene expression, viability, and fertility are unaffected.

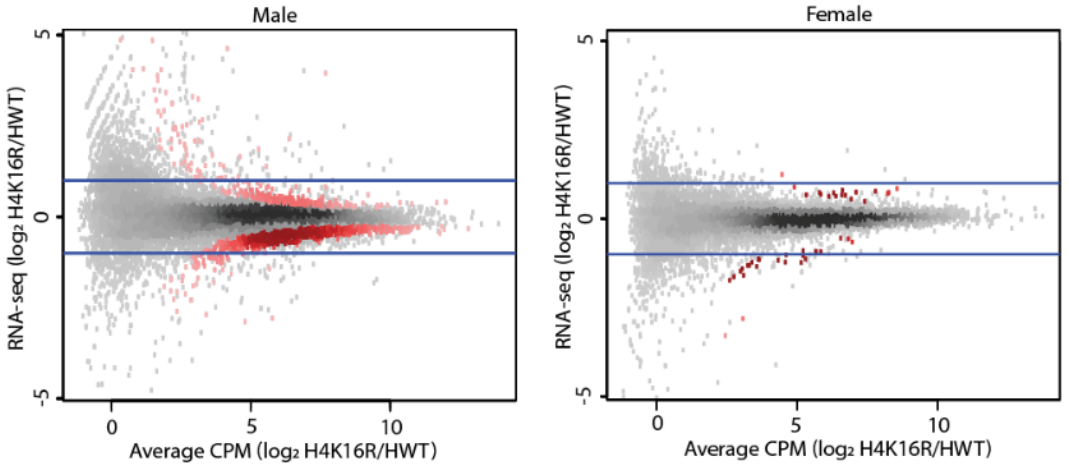
Figure 2.14. H4K16 promotes hyper-expression of the *Drosophila* male X chromosome. A) Table of observed HWT and H4K16R adult females and males where first instar larvae of each genotype were isolated from their wild-type siblings and mono-cultured in aliquots of 50 larvae per vial (see supplementary materials for crosses and complete genotypes: Rows 1 and 2) zygotic, replication-dependent HWT and H4K16R; Rows 3 and 4) maternal/zygotic, replication-dependent HWT and H4K16R; and Rows 5 and 6) zygotic, replication-dependent and replication-independent HWT and H4K16R (left; Chi squared comparisons performed against the male to female ratio of zygotic, HWT, $p<0.01$;). Percentage of viable male (grey) and female (black) adults for H4K16R and HWT (right). B) Heatscatter plot of the H4K16R/HWT ratio of RNA-seq signal from wing imaginal discs. Statistically different transcripts between H4K16R and HWT males (left panel) and H4K16R and HWT females (right panel) are indicated in red ($p<0.05$). Blue lines indicate a two-fold change. C) Box plot of RNA-seq signal from autosomes and ChrX after MSL2 or MOF knockdown in male S2 cells (Zhang et al. 2010) and in H4K16R/HWT male and female wing discs on autosomes (Auto) and ChrX. D) Average enrichment of modENCODE H4K16ac signal from male third instar larvae at 10kb windows of significantly ($p<0.05$) decreased (dec) or increased (inc) transcript expression between H4K16R and HWT males on ChrX and autosomes (Auto) or at all 10kb windows (GSE49497) (Celniker et al. 2009).

A

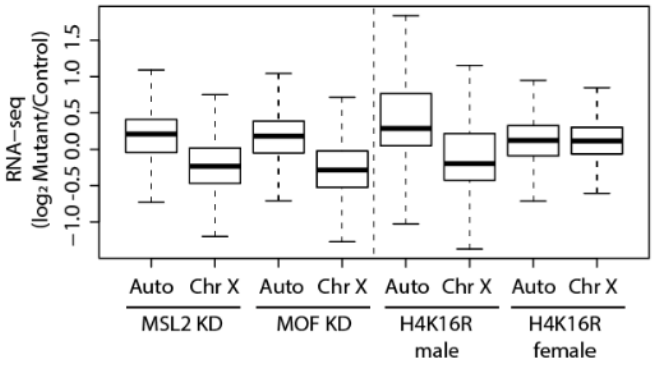
	Genotype	Total # of larvae	Female Observed Adults	Male Observed Adults	M:F $\chi 2$
Zyg	ΔHisC; HWT/+	450	141	154	-
	ΔHisC; H4K16R/+	550	188	87	<0.0001
Mat, Zyg	ΔHisC; HWT/+	300	116	110	n.s.
	ΔHisC; H4K16R/+	250	97	9	<0.0001
Zyg, His4rΔ	ΔHisC; HWT, His4rΔ/His4rΔ	400	157	148	n.s.
	ΔHisC; H4K16R, His4rΔ/His4rΔ	400	140	0	<0.0001



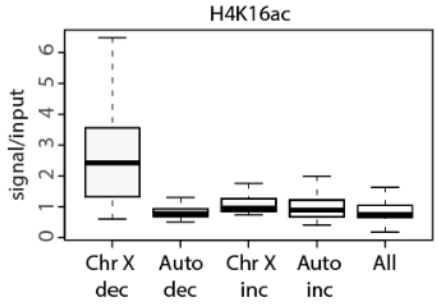
B



C



D



H4K16 promotes early replication of the *Drosophila* male X Chromosome

We next profiled RT in replication dependent *H4K16R* and control male and female wing discs. When considering the major chromosome scaffolds using overlapping 100kb windows, we observed very few significant replication changes between *H4K16R* and control in either females or males (0.04% and 1%, respectively) (Fig 2.15; Fig 2.16). These data indicate that H4K16 is not globally required for maintenance of RT in flies. However, when we considered 100kb windows only from the X Chromosome, we observed that the normally earlier replication of the male X relative to the female X was largely abrogated in *H4K16R* mutants ($p<0.05$; Fig 2.17A,B), suggesting that H4K16ac promotes early replication of the male X Chromosome.

To evaluate this effect more thoroughly, we assigned RT values to 10kb windows across the genome using significance thresholds as for the *H3K9R* RT data. We identified 57 individual 10kb windows in *H4K16R* males with delayed RT, and most (78%) of these were located on the X (Fig 2.17C; Fig 2.18A). We identified 92 10kb windows in *H4K16R* males with advanced RT, with most (94%) located on the autosomes (including 61 windows on Chr 3R) (Fig 2.17C; Fig 2.18A). Windows from the *H4K16R* male X with delayed replication are enriched for H4K16ac in a wild type genome, whereas those advanced windows (on the X or autosomes) or delayed windows on autosomes are not ($p < 0.05$; Fig 2.17D; Fig 2.18B). These data suggest that delayed replication in *H4K16R* males is a direct result of the *H4K16R* mutation, while regions of advanced replication may occur indirectly.

We were concerned that the small number of windows with a RT change identified using our significance cutoffs was masking a more general effect as many X Chromosome replication delays in *H4K16R* males might be less than a $\log_2$ fold change of 0.1, and

therefore not scored as significant. Indeed, chromosome-wide RT of the X in control males is advanced less than a $\log_2$ fold change of 0.1 relative to control females (Fig 2.17A,B) in accordance with previous analyses (e.g. a change of ~0.1 as described by (Schwaiger et al. 2009; Lubelsky et al. 2014)). Therefore, we analyzed replication in *H4K16R* males by assessing, as a group, all 10 kb windows located on an individual chromosome (X and 4) or an individual chromosome arm (2L, 2R, 3L, and 3R). Using this approach, replication of the X was found to be significantly delayed in *H4K16R* males relative to control males (Fig 2.17E). No such effect was observed for the autosomes, consistent with a specific role for H4K16ac in promoting early replication of the male X. The small, heterochromatic 4th Chromosome replicated earlier in both *H4K16R* males (Fig 2.17E) and females (Fig 2.18C) which may result from an H4K16 function that is independent of X Chromosome dosage compensation. We conclude that H4K16 promotes early replication of the male X in *Drosophila.*

Figure 2.15. Replication timing profile for *H4K16R* females and control. LOESS regression line applied to S/G1 (log_2) averaged replicates from *HWT* female (yellow) and *H4K16R* female (black) plotted across all major chromosome scaffolds at 100kb windows with a 10kb slide.

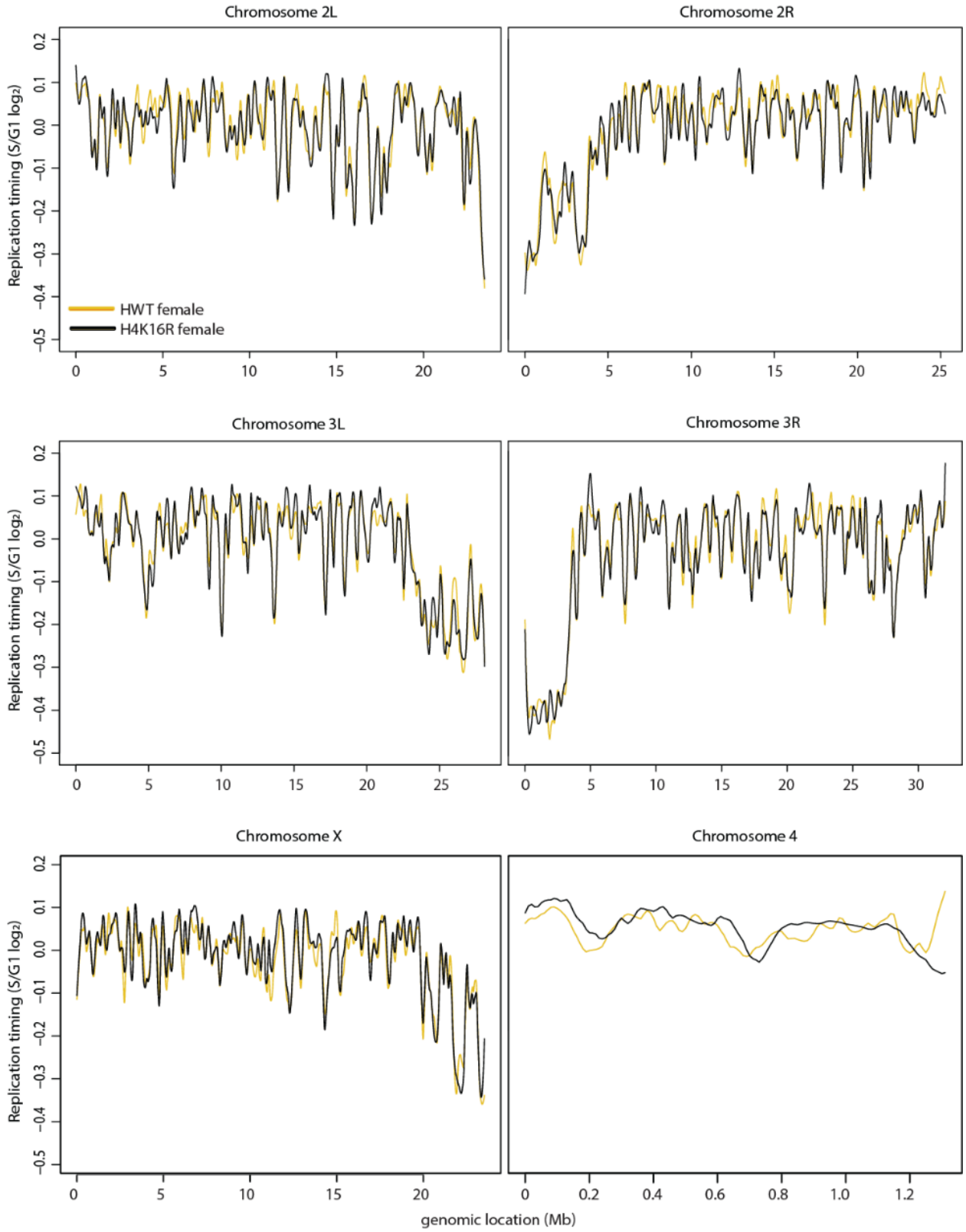

Chromosome 2L
Chromosome 2R
Chromosome 3L
Chromosome 3R
Chromosome X
Chromosome 4
Replication timing (S/G1 log2)
HWT female
H4K16R female
genomic location (Mb)

Figure 2.16. Replication timing profile for *H4K16R* males and control. LOESS regression line applied to S/G1 (log_2) averaged replicates from *HWT* male (maroon) and *H4K16R* male (blue) plotted across all major chromosome scaffolds at 100kb windows with a 10kb slide.

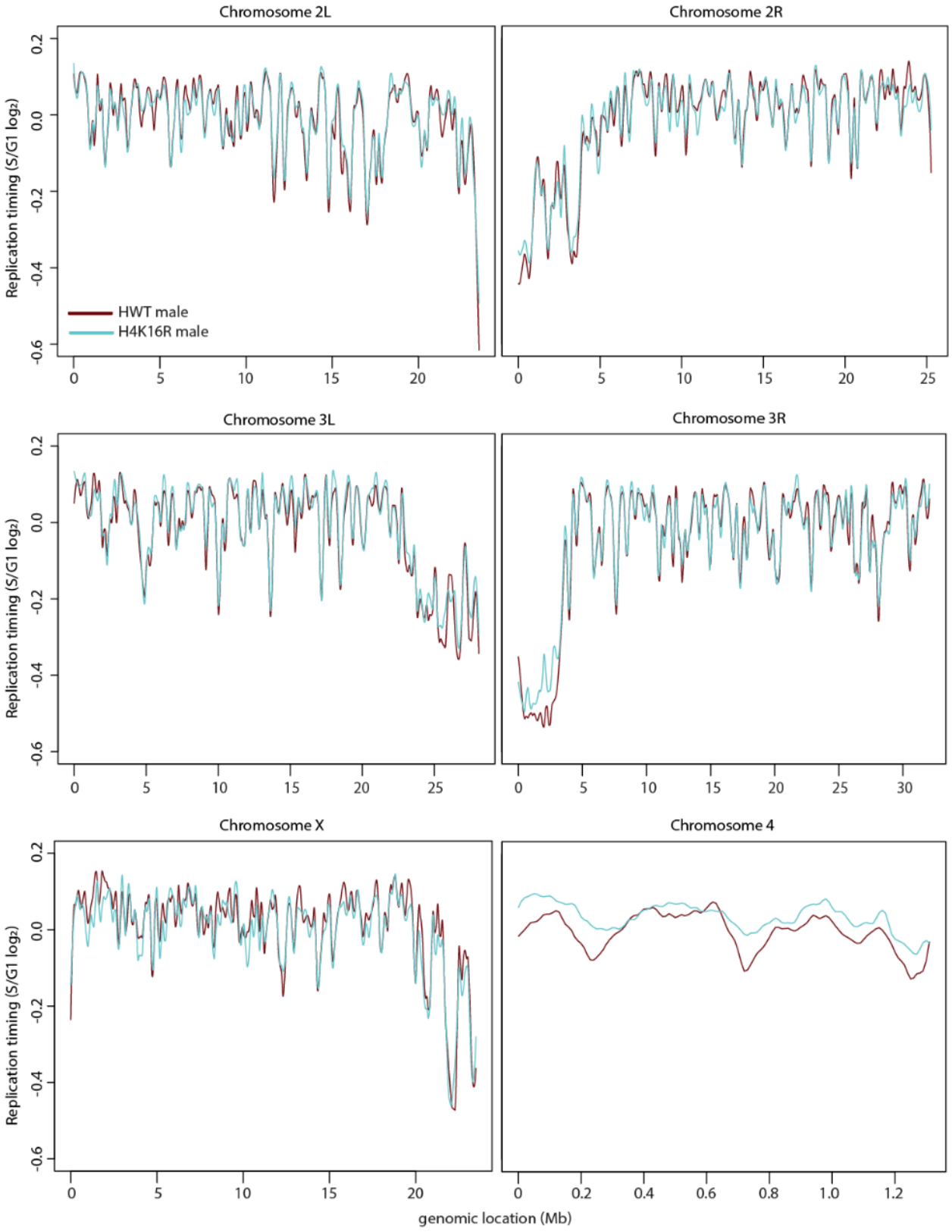

Chromosome 2L
Chromosome 2R
Chromosome 3L
Chromosome 3R
Chromosome X
Chromosome 4
Replication timing (S/G1 log2)
genomic location (Mb)
HWT male
H4K16R male

Figure 2.17. *H4K16R* mutation reduces gene expression and delays replication of the male X. A) Boxplot of *HWT* male/female and *H4K16R* male/female ratios of RT values (log_2 S/G1) on Chr X. B) LOESS regression line applied to log_2 S/G1 averaged replicates from *HWT* female (yellow) and *HWT* male (maroon) and *H4K16R* female (black) and *HWT* male (blue) plotted across Chr X (100kb windows, 10kb slide). Note that the male X Chromosome generally replicates earlier in *HWT*, but not in *H4K16R* mutants. C) Histogram of 10kb windows with advanced (red) or delayed (blue) RT between *H4K16R* and *HWT* males on major chromosome scaffolds ($p<0.05$; absolute log_2 fold change>0.1; limma). D) Average enrichment of modENCODE H4K16ac signal from male third instar larvae at 10kb windows of delayed (del) or advanced (adv) replication between *H4K16R* and *HWT* males on Chr X and autosomes (Auto) or at all 10kb windows (GSE49497) (Celniker et al. 2009). E) Box plot of the *H4K16R*/*HWT* ratio of male RT values (log_2 S/G1) on all major chromosome scaffolds. F) Box plot of the *H4K16R*/*HWT* ratio of male RT values (log_2 S/G1) at 10kb windows of decreased or increased RNA-seq signal on Chr X or autosomes (Auto) ($p<0.05$). G) Box plot of the *H4K16R*/*HWT* ratio of male RNA-seq signal at 10kb windows of delayed or advanced RT ($p<0.05$). H) Heatscatter plot of the *H4K16R*/*HWT* ratio of male RT values (log_2 S/G1) plotted versus the *H4K16R*/*HWT* ratio of male RNA-seq signal at all 10kb windows across the autosomes (left) and Chr X (right). RNA-seq differences were determined based on the transcript with the lowest p-value across the 10kb window. The percentage of 10kb windows present in each quadrant is indicated.

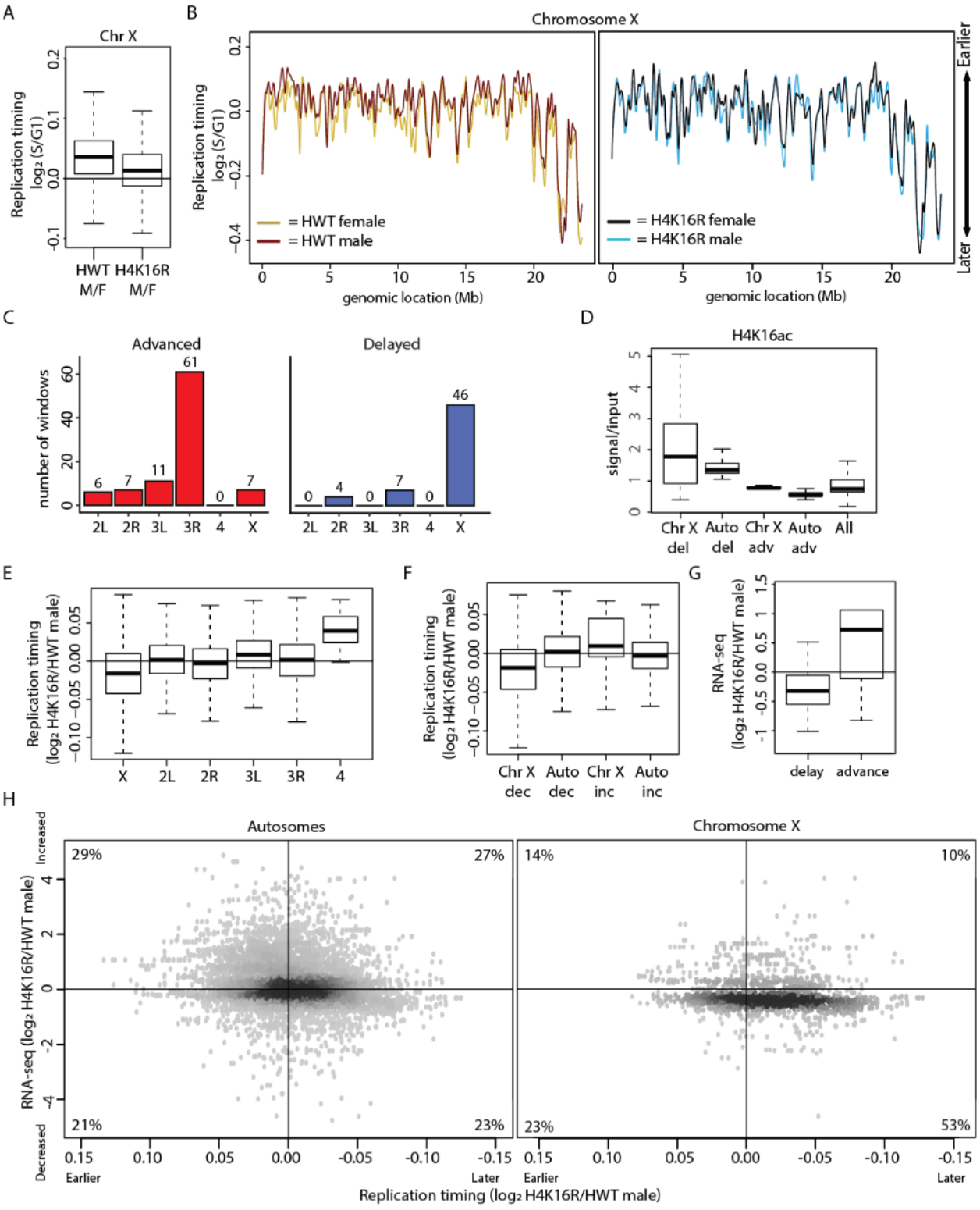
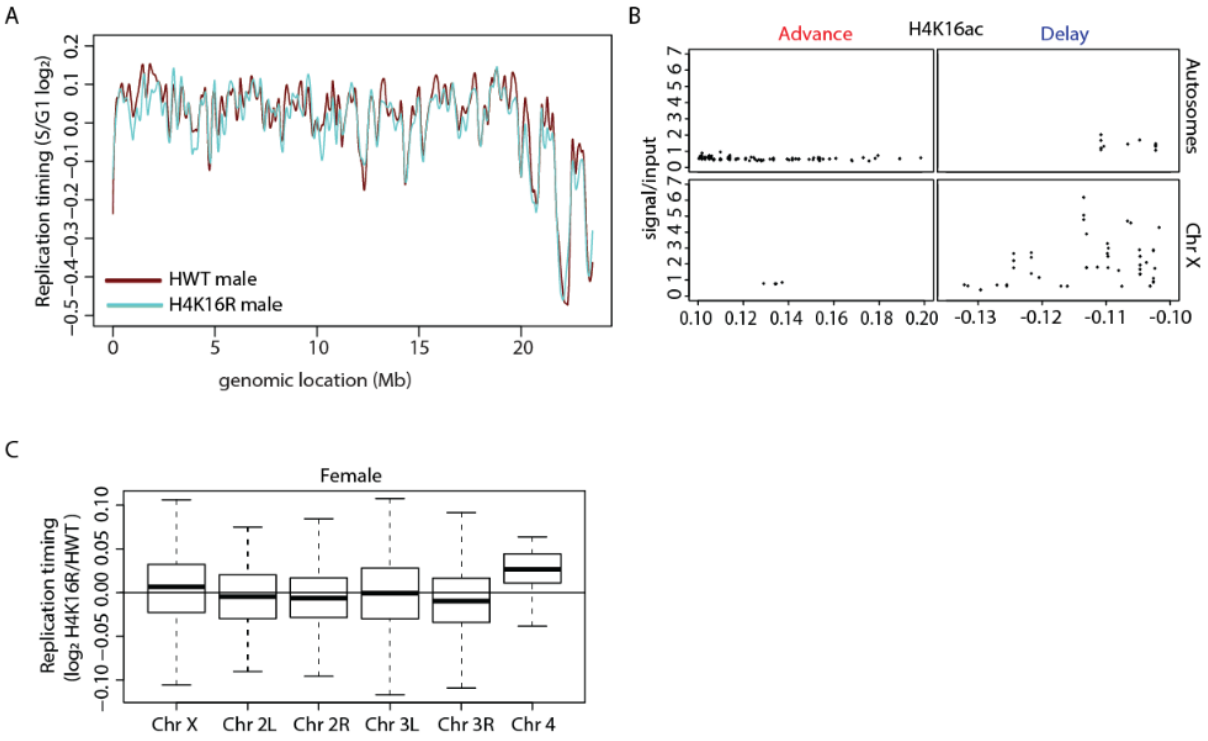

A
Chr X
Replication timing log2 (S/G1)
HWT M/F
H4K16R M/F
B
Chromosome X
Replication timing log2 (S/G1)
= HWT female
= HWT male
= H4K16R female
= H4K16R male
genomic location (Mb)
Earlier
Later
C
Advanced
Delayed
number of windows
2L 2R 3L 3R 4 X
D
H4K16ac
signal/input
Chr X del
Auto del
Chr X adv
Auto adv
All
E
Replication timing (log2 H4K16R/HWT male)
X 2L 2R 3L 3R 4
F
Chr X dec
Auto dec
Chr X inc
Auto inc
G
RNA-seq (log2 H4K16R/HWT male)
delay
advance
H
Autosomes
Chromosome X
RNA-seq (log2 H4K16R/HWT male)
Increased
Decreased
29%
27%
21%
23%
14%
10%
23%
53%
Earlier
Later
Replication timing (log2 H4K16R/HWT male)

Figure 2.18. Characterization of altered replication timing in *H4K16R* mutants. A) LOESS regression line applied to S/G1 (log_2) averaged replicates from *HWT* male (maroon) and *H4K16R* male (blue) plotted across Chromosome X at 100kb windows with a 10kb slide. B) Correlation analysis of *H4K16R*/*HWT* male log_2 RT fold change versus the average enrichment of H4K16ac signal at 10kb windows on autosomes (top) and Chr X (bottom) with significantly at advanced (left) and delayed (right) replication. ChIP-seq enrichment was determined from modENCODE datasets from male wild-type whole 3rd instar larvae. C) Box plot of the *H4K16R*/*HWT* female ratio of normalized replication timing values (S/G1 (log_2)) on all major chromosome scaffolds at 100kb windows with a 10kb slide.



***H4K16R* mutation concurrently reduces gene expression and delays replication of the male X**

To explore the relationship between gene expression and RT of individual windows on the male X, we identified 10kb windows containing a differentially expressed transcript in *H4K16R* males relative to controls and determined whether these windows also displayed altered replication. For the X, we found a significant correlation between decreased gene expression and later replication in *H4K16R* males, as well as a correlation between increased gene expression and earlier replication ($p<0.05$; Fig 2.17F). No such correlation exists for windows on the autosomes (Fig 2.17F). These correlations hold when we consider the converse relationship: 10kb windows in *H4K16R* males with significantly delayed replication have decreased gene expression whereas 10kb windows with significantly advanced replication have increased gene expression ($p<0.05$; Fig 2.17G). These data indicate that changes in RT correlate with changes in gene expression for the male X in *H4K16R* males.

Furthermore, scatterplots comparing transcription and replication between *H4K16R* males and controls resulted in different distributions. We found that 53% of all 10kb windows on Chromosome X had both lower gene expression and later replication in *H4K16R* males compared to 23% on autosomes (Fig 2.17H). A two-dimensional Peacock test (Peacock 1983) revealed that the distributions resulting from simultaneous comparison of the $\log_2$ fold change of the *H4K16R*/control ratio of transcript abundance and RT for all 10kb windows on the X and the autosomes are statistically different ($P < 5.9*10^{-317}$). These data indicate that H4K16 promotes the strong correlation between elevated gene expression and early replication of the *Drosophila* male X Chromosome.

Discussion

Studies of animal cells have revealed strong, genome-wide correlations between early replication of highly transcribed, accessible chromatin and late replication of lowly transcribed, inaccessible chromatin (Bell et al. 2010; Eaton et al. 2011; Lubelsky et al. 2014). Here we explored potential causal relationships underlying these correlations by combining perturbation of chromatin structure using two different histone mutations (*H3K9R* and *H4K16R*) with genome-wide RT data.

We found that the male X Chromosome of *H4K16R* mutants experiences both reduced transcription and delayed RT, consistent with previous studies showing a strong correlation between transcriptional activity and RT (Aggarwal and Calvi 2004; MacAlpine et al. 2004; Schwaiger et al. 2009; Liu et al. 2012; Lubelsky et al. 2014). However, we also show that correlations between transcription and RT can be uncoupled: active transposon expression in *H3K9R* mutants was not accompanied by earlier replication of most pericentric heterochromatin. Thus, activation of transcription does not always result in earlier replication. In addition, our analysis of *H4K16R* mutants show for the first time that H4K16 is required for proper dosage compensated expression of the *Drosophila* male X Chromosome, as predicted by previous studies of factors that acetylate H4K16 (Hilfiker et al. 1997; Smith et al. 2000; Kuroda et al. 2016). The changes in autosomal gene expression we observed in *H4K16R* males are likely a secondary consequence of wholesale changes in gene expression on the X, as transcription factors are encoded on the X Chromosome.

HP1a binding to H3K9me is a defining feature of constitutive heterochromatin and is thought to be critical for most if not all aspects of heterochromatin function (Canzio et al. 2011; Larson et al. 2017). Therefore, we were surprised to find that pericentric

heterochromatin generally remained late replicating relative to the euchromatic chromosome arms in *H3K9R* mutants. Indeed, despite decreased nucleosome density and loss of HP1a, replication at ~97% of the genome remained unchanged in *H3K9R* mutants. Studies of the onset of late replication in the early fly embryo show that chromatin condensation and late replication of pericentric heterochromatin occur prior to H3K9me and HP1a recruitment, indicating that these two features of heterochromatin are not always required for late replication (Shermoen et al. 2010; Yuan and O'Farrell 2016). In fact, we observe a DAPI-bright heterochromatic chromocenter in *H3K9R* diploid nuclei, similar to that of wild type nuclei, that colocalizes with late patterned replication foci. Thus, pericentric heterochromatin retains many of its hallmarks despite loss of H3K9me and HP1a, suggesting additional features define heterochromatin function.

Our data support a model in which compartmentalization of euchromatin and heterochromatin into different nuclear compartments is not disrupted by the loss of H3K9 modification. Furthermore, the arms and pericentric regions of *Drosophila* chromosomes may correspond, respectively, to the largely euchromatic compartment “A” and heterochromatic compartment “B” previously identified in human cells (Lieberman-Aiden et al. 2009). Accordingly, factors other than HP1a may remain associated with compartment “B” in *H3K9R* mutants, preventing large-scale advancement of RT at the pericentromere. One such factor could be Rif1, which is required for the onset of late replication of heterochromatin during early fly embryogenesis as well as for late replication in other species (Peace et al. 2014; Foti et al. 2016; Seller and O’Farrell 2018).

Nevertheless, we found reproducible and significant RT changes in *H3K9R* pericentric heterochromatin. By carefully analyzing features of these altered RT domains, we

conclude that accessible chromatin does not invariably result in early replication, although early replication may require accessible chromatin. We propose that *H3K9R* mutation alters RT by disrupting local chromatin accessibility without affecting overall compartmentalization of heterochromatin (Larson et al. 2017; Strom et al. 2017). The events that function within accessible chromatin to dictate RT could include origin specification or origin activation. Origins of replication are licensed during G1 phase by the activity of origin specification factors (e.g. the origin recognition complex; ORC), and during S phase DNA replication initiates at a subset of licensed origins (Bell and Stillman 1992). Certain models describing temporal programs of replication initiation posit a stochastic process in which a higher density of licensed origins in accessible euchromatin increases the probability of replication initiation compared to inaccessible heterochromatin (Rhind et al. 2010; Das et al. 2015; Miotto et al. 2016). These models are consistent with observations that ORC complexes are most abundant where chromatin accessibility is also high (MacAlpine et al. 2010; Lubelsky et al. 2014; Miotto et al. 2016). However, other factors function within the licensed origin landscape to either promote or inhibit origin activation (Foti et al. 2016). Thus, changes in either origin licensing or activation could demarcate domains that advance replication within the permissive open chromatin environment created by the *H3K9R* mutation.

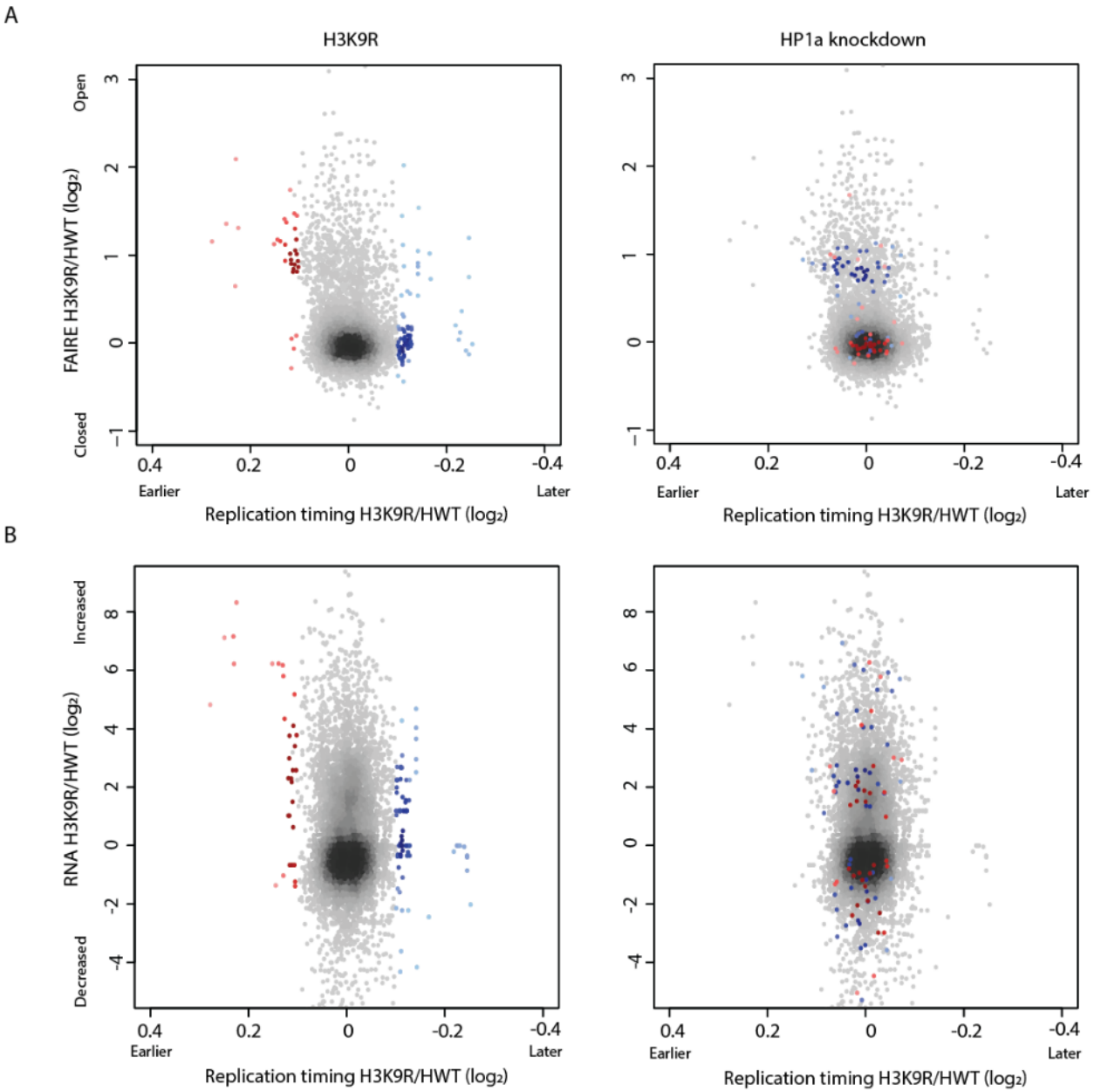
We also found that delayed replication domains are largely independent of altered chromatin accessibility or transcriptional changes in *H3K9R* mutants. We hypothesize that elevated accessibility of pericentric heterochromatin in *H3K9R* mutants functions as a "sink" for limiting replication factors, resulting in delayed replication of domains along chromatin arms as proposed for other replication factors (Foti et al. 2016). Another possible explanation

for delayed replication in *H3K9R* mutants is the relocalization of HP1a to chromosome arms (Penke et al 2016). Previous work has shown that tethering HP1a to a euchromatic domain delays RT (Pokholkova et al. 2014). However, we did not observe a strong correlation between 10kb windows that gain HP1a and those that delay replication, perhaps because the amount of HP1a relocalization in *H3K9R* mutants was below a threshold necessary to affect replication. We note that domains of altered RT in *H3K9R* mutants do not match those previously identified after HP1a knockdown in *Drosophila* cultured cells (Fig 2.19) (Schwaiger et al. 2010), potentially due to H3K9-independent functions of HP1a or to the exclusion of repetitive DNA from the microarray based assay used in the previous study.

In summary, our study shows that correlations among chromatin configuration, transcription, and RT in animal cells can be mechanistically separated by mutation of specific histone residues, indicating modes of control for replication initiation that are independent of these features of animal genome structure and activity.

Figure 2.19. Domains of altered replication in *H3K9R* mutants do not overlap those identified after HP1a knockdown.

A-B) Heatscatter plot of the *H3K9R*/*HWT* ratio of normalized replication timing values (S/G1 (log_2)) plotted versus the *H3K9R*/*HWT* ratio of normalized FAIRE (A) or RNA-seq (B) signal at Hidden Markov Model determined replication domains identified by (Schwaiger et al. 2010). Significantly advanced windows are indicated in red and significantly delayed windows are indicated in blue. Domains identified as significantly altered in *H3K9R* mutants are shown in the left panels and domains identified by Schwaiger et al. using BrdU ChIP coupled with microarrays are shown in the right panels. Differences between the two datasets could be due to H3K9-independent roles for HP1 or tissue specific differences (3^{rd} instar imaginal wing disc vs. embryo derived Kc cells). We speculate that regions of advanced replication timing identified in our study that were not identified in Schwaiger et al. were due to the removal of repetitive DNA sequences in the microarray designed used previously.


A
H3K9R
HP1a knockdown
Open
Closed
FAIRE H3K9R/HWT (log2)
3
2
1
0
−1
0.4
0.2
0
-0.2
-0.4
Earlier
Later
Replication timing H3K9R/HWT (log2)
B
Increased
Decreased
RNA H3K9R/HWT (log2)
8
6
4
2
0
-2
-4

Acknowledgements

This work was supported by NIH Grants R01-GM124201 to R.J.D.; F31-GM115194 to T.J.R.P; R01-GM104097 to D.M.M.; R35-GM128851 to D.J.M.; and R01-GM129132 to A.G.M. In addition, R.L.A. and T.J.R.P. were supported in part by an NIH predoctoral training grant T32-GM007092. We thank the UNC Flow Cytometry and High Throughput Sequencing Core Facilities, supported in part by P30 CA016086 Cancer Center Core Support Grant to the UNC Lineberger Comprehensive Cancer Center. FACS results reported in this publication were supported in part by the North Carolina Biotechnology Center Institutional Support Grant 2012-IDG-1006. We thank Janet Dow and Corbin Jones for data acquisition and analysis advice. We thank Spencer Nystrom for critical reading of the manuscript and Agnes Ezekwesili and Esther Kwon for assistance with larval dissection.

CHAPTER 3- H3K9 PROMOTES UNDER-REPLICATION OF PERICENTROMERIC HETEROCHROMATIN IN *DROSOPHILA* SALIVARY GLAND POLYTENE CHROMOSOMES[2]

Introduction

Proper genome duplication is essential for normal development and tissue homeostasis. In diploid cells, genome duplication and cell proliferation occur via canonical G1→S→G2→M cell cycles in which origins of replication are specified during G1 phase, DNA replication occurs during S phase, and chromosome segregation and cell division occurs during M phase (Bell 2017). Many diploid organisms, including humans, contain tissues composed of cell types with a polyploid genome, a phenomenon called endopolyploidy that serves as a developmental strategy for tissue growth and generating cells with high biosynthetic capacity (Lee et al. 2009; Zielke et al. 2013; Orr-Weaver 2015). Although endopolyploidy is a common feature of normal development in both plants and animals, polyploidy can also result from mis-regulation of the canonical diploid cell cycle and is commonly associated with human disease (Lee et al. 2009; Fox and Duronio 2013). Therefore, determining mechanisms that regulate polyploid cell cycles is important for understanding both normal and pathological development.

[2] [2] This chapter previously appeared as an article in Genes. The original citation is as follows: Armstrong, R.L., Penke, T.J.R., Chao, S.K., Gentile, G.M., Strahl, B.D., Matera, A.G., McKay, D.J., and Duronio, R.J., 2018. H3K9 promotes under-replication of pericentromeric heterochromatin in Drosophila salivary gland polytene chromosomes. *Genes.* Accepted.

Polyploidy often arises from endoreplication, a cell cycle in which repetitive rounds of DNA replication occur without intervening mitosis and cell division (Edgar and Orr-Weaver 2001; Lilly and Duronio 2005; Lee et al. 2009; Fox and Duronio 2013; Zielke et al. 2013; Orr-Weaver 2015; Hua and Orr-Weaver 2017). The giant polytene chromosomes of the polyploid cells of the *Drosophila* larval salivary gland have long served as a model experimental tissue for understanding endoreplication (Lilly and Duronio 2005; Zielke et al. 2011). During late embryonic and larval development, approximately ten endoreplication cycles yield a final ploidy of ~1350C in salivary gland cells (Hammond and Laird 1985; Zielke et al. 2011). These polyploid cells use the same *trans*-acting factors as diploid cells to control DNA replication initiation (Zielke et al. 2011), which occurs stochastically from many origins of replication throughout S phase, yielding reproducibly earlier and later replicating domains (Kolesnikova et al. 2018). However, replication is not uniform across the salivary gland polyploid genome as it is in diploid cells. Whereas the earlier replicating regions of the genome are duplicated each endocycle, the latest replicating regions are not replicated each endocycle, resulting in under-replicated domains (Hammond and Laird 1985; Spradling and Orr-Weaver 1987). Stalled replication forks at these under-replicated domains cause DNA damage, resulting in deletions that contribute to copy number reduction (< 1350C) (Andreyeva et al. 2008; Yarosh and Spradling 2014). Under-replicated domains also occur in mammalian polyploid cells (Hannibal et al. 2014) and share characteristics of mammalian diploid cell fragile sites (Hua and Orr-Weaver 2017). Thus, the study of the origin and properties of under-replicated domains in polyploid genomes will help us understand general features of genome organization and stability.

The mechanistic basis for under-replication is not completely understood, but recent studies suggest contributions from regulating origin firing and replication fork progression (Sher et al. 2012; Hua and Orr-Weaver 2017). In both diploid and polyploid cells the timing of DNA replication initiation during S phase correlates with chromatin organization: transcriptionally active, accessible euchromatin generally replicates earlier during S phase whereas transcriptionally repressive, inaccessible heterochromatin generally replicates later (MacAlpine et al. 2004; Bell et al. 2010; Eaton et al. 2011; Lubelsky et al. 2014; Pokholkova et al. 2015; Prioleau and MacAlpine 2016; Armstrong et al. 2018). This differential replication timing is regulated, in part, by non-uniform distribution of origins of replication throughout the genome. Euchromatic regions of the genome have a higher density of origins relative to heterochromatic regions, resulting in a higher probability of DNA replication initiation in euchromatin relative to heterochromatin (Rhind et al. 2010; Sher et al. 2012; Das et al. 2015; Miotto et al. 2016). However, a paucity of origins is an insufficient explanation for under-replication in polyploid cells because some regions of the genome that constitutively lack origins are not under-replicated in all *Drosophila* polyploid cell types (Sher et al. 2012; Hua et al. 2018). Rather, reduced origin density coupled with inhibition of replication fork progression contribute to under-replication within polyploid genomes (Sher et al. 2012; Kolesnikova et al. 2013; Hua et al. 2018; Munden et al. 2018). The latest replicating regions located within pericentric heterochromatin experience the greatest degree of under-replication.

Several heterochromatin-associated proteins contribute to under-replication in the *Drosophila* salivary gland. Heterochromatin Protein 1a (HP1a) binds di- and tri-methylated H3K9, which is enriched in pericentric heterochromatin, and facilitates heterochromatin

formation through multimerization of HP1a molecules and recruitment of other heterochromatin-associated factors (Canzio et al. 2011; Larson et al. 2017; Strom et al. 2017). The SuUR (Suppressor of Under-Replication) protein, a SNF2-like component of silent chromatin in both diploid and polyploid cells (Makunin et al. 2002), localizes to late replicating heterochromatin and inhibits DNA replication fork progression to promote under-replication (Belyaeva et al. 1998; Makunin et al. 2002; Kolesnikova et al. 2013; Nordman et al. 2014). HP1a and SuUR recruitment to *Drosophila* salivary gland chromosomes are interdependent on one another: both the absence and over-expression of HP1a disrupt SuUR chromatin binding, and over-expression of SuUR results in mis-localization of HP1a to ectopic SuUR sites (Pindyurin et al. 2008). Furthermore, tethering either SuUR or HP1a to earlier replicating regions of salivary gland polytene chromosomes is sufficient to delay replication but not to induce under-replication (Pokholkova et al. 2015). Rif1 (Rap1-Interacting Factor 1) and the linker histone H1 both directly interact with SuUR and are required for under-replication (Andreyeva et al. 2017; Munden et al. 2018). H1 functions upstream of SuUR and is required for SuUR chromatin binding (Andreyeva et al. 2017). Furthermore, although Rif1 directly regulates replication fork progression in a SuUR-dependent manner, SuUR localization to under-replicated regions is independent of Rif1 (Munden et al. 2018).

In contrast to our current understanding of the contributions of *trans*-acting factors, the roles in endoreplication and under-replication of individual histone tail residues that impact chromatin organization have not been determined. The ninth lysine on histone H3 (H3K9) and the sixteenth lysine on histone H4 (H4K16) have been implicated in promoting S phase progression through studies that modulate factors that catalyze (writers) or bind

(readers) the post-translational modifications of these residues (De Lucia et al. 2005; Peng and Karpen 2009; Li et al. 2010; Zhao et al. 2013). Notwithstanding their importance, these studies cannot directly determine whether H3K9 and H4K16 themselves regulate S phase progression, as histone modifying enzymes also modify non-histone substrates (Glozak et al. 2005; Huang and Berger 2008; Sims and Reinberg 2008). To address this issue, we employed a strategy in *Drosophila* to generate histone mutant genotypes, an approach that is not currently feasible in other animals due to the large number of replication-dependent histone genes. The strategy involves deleting the endogenous wild type histone genes and replacing them with transgenic copies encoding an amino acid substitution that prevents post-translational modification of a particular histone residue (Günesdogan et al. 2010; McKay et al. 2015).

We recently demonstrated that, in contrast to mutation of H3K9 writers, readers, and erasers, *H3K9R* mutant *Drosophila* diploid wing imaginal discs have only a modestly reduced S phase index, suggesting a small role played by H3K9 in canonical S phase progression (Armstrong et al. 2018). We observed a similarly modest effect on S phase progression in *H4K16R* wing imaginal disc cells (Armstrong and Duronio). Here, we utilize *H3K9R* and *H4K16R* mutations to probe the role of heterochromatin and euchromatin, respectively, in cell cycle phasing, DNA replication timing, and under-replication in *Drosophila* salivary gland polytene chromosomes. We demonstrate that H3K9 regulates endoreplication whereas H4K16 is largely dispensable. Furthermore, we demonstrate that H3K9 promotes under-replication of pericentric heterochromatin whereas under-replication along chromosome arms is H3K9-independent.

Materials and Methods

Drosophila larval culturing

All fly stocks were maintained on standard corn medium and crossing schemes to generate engineered replication-dependent histone genotypes were performed as in (Armstrong et al. 2018). Fifty GFP-positive Histone Wild Type (*HWT*), *H3K9R*, or *H4K16R* larvae were cultured independently of their phenotypically wild type, GFP-negative siblings by manually moving first-instar larvae into vials of standard corn medium and allowing them to develop until third instar larvae. Note that only the replication-dependent *H3* and *H4* histone genes were mutant in this study.

Salivary gland polytene chromosome immunofluorescence

Pre-wandering third-instar larvae were staged using the following criteria: crawling on top of the media, not displaying wandering behavior on vial edges, and no longer eating. Salivary glands were dissected from pre-wandering third-instar larvae in 1xPBT (0.1% Triton X-100 in PBS, pH 7.5). Glands were permeabilized and fixed in the following solutions: 1) 2 minutes in (3.7% Paraformaldehyde, 1xPBT), 2) 2 minutes in (3.7% Paraformaldehyde, 50% Acetic Acid), and moved to 3) 1:2:3 Lactic Acid: dH_20: Acetic Acid on a siliconized coverslip. Spread polytene chromosomes were flash frozen in liquid N_2 and stored in 1xPBT until all slides were completed. Polytene chromosome spreads were incubated in Image-iT FX Signal Enhancer (Thermo Fisher) for 30 minutes at room temperature, stained with 1:500 mouse anti-PCNA (Abcam; ab29) and 1:1500 mouse anti-HP1a (DSHB; C1A9) in 1xPBT overnight at 4 degrees, treated with 0.2ug/mL DAPI for 5 minutes at room temperature, mounted in ProLong Gold antifade reagent, and imaged on a Leica confocal microscope.

PCNA patterns were staged according to the following criteria: ER) dim PCNA signal across chromosome arms with few gaps between bands of PCNA signal; E-MR) bright PCNA signal across chromosome arms with few, distinct gaps between bands of PCNA signal; M-LR) thick bands of PCNA signal across chromosome arms with large gaps between bands of PCNA signal and bright chromocenter PCNA signal; LR) thin bands of PCNA signal across chromosome arms with large gaps between bands of PCNA signal and chromocenter PCNA signal; VL-R) dim, sparse PCNA signal primarily at the chromocenter; NR) no PCNA signal.

Sample preparation for genome sequencing

Salivary glands from female, wandering third-instar larvae were isolated and flash frozen in liquid N_2 until all samples were collected. Nuclei were isolated from replicates of 25 salivary glands and sonicated with a Branson Sonifier 450 to an average fragment size distribution of 500-1000bp. Sonicated samples were treated with 100 µg/mL RNaseA at 37°C for 1 hour and with 200 µg/mL proteinase K at 37°C for 2 hours. Genomic DNA was phenol chloroform extracted and stored at -80°C prior to library preparation. Libraries were prepared from 20 ng of genomic DNA with the ThruPLEX DNA-seq kit (Rubicon Genomics) and sequenced on an Illumina HiSeq2500 at the UNC High-Throughput Sequencing Core Facility. Sequencing data can be obtained using GEO accession number GSE125505.

Bioinformatics

Paired-end 100bp reads were trimmed using Trimmomatic (v0.36) with LEADING:30 and TRAILING:30 parameters and aligned to the dm6 reference genome (release 6.04) using Bowtie2 (v2.3.2) default parameters. Reads with a MAPQ score greater

than 10 were retained with SAMtools (v1.6) which removes reads with low confidence mappability that often include simple repeats. Thus, reads within heterochromatin map uniquely; see (Penke et al. 2016; Armstrong et al. 2018; Penke et al. 2018). BEDTools coverage (v2.25.0) was used to quantify the number of reads mapping to 10kb windows tiled across the genome, with results normalized to read depth (Quinlan and Hall 2010). CNVnator 0.3.3 was used for detection of under-replicated sequences using a bin size of 1000 (Abyzov et al. 2011). Under-replicated domains were called if they were: 1) detected by CNVnator and 2) greater than 10kb in size. Pericentromeric and chromosome arm boundaries were defined by high levels of H3K9me2 enrichment as in (Riddle et al. 2011; Hoskins et al. 2015). A p value of 0.001 was used as the statistical significance cutoff for all bioinformatic analyses. modENCODE H3K9me2 ChIP-seq data from whole third instar larvae (GEO accession number GSE47260) was used for analyses and LOESS regression lines were generated with span = 0.02.

Results

We used our histone gene replacement platform to generate *H3K9R* and *H4K16R* mutant larvae and determine whether replication-dependent H3K9 and H4K16 are necessary for endoreplication in the *Drosophila* salivary gland. The approximately ten salivary gland endoreplication cycles occur from 7-96 hours after egg deposition and are completed by the wandering third-instar larval stage of development (Hammond and Laird 1985; Zielke et al. 2011). To visualize nuclei actively undergoing DNA replication, we prepared salivary gland polytene chromosome spreads from pre-wandering third-instar larvae and stained them with an antibody against Proliferating Cell Nuclear Antigen (PCNA), as done previously

(Andreyeva et al. 2017; Kolesnikova et al. 2018). PCNA travels with the replisome and functions as a DNA polymerase processivity factor (Slade 2018), and thus serves as a cytological marker of active replication forks (Figure 3.1A) (Grant et al. 2018). Consequently, PCNA-positive polytene chromosome spreads are undergoing endo-S phase whereas PCNA-negative polytene chromosome spreads are not (i.e. G phase) (Figure 3.1A).

As a general assessment of endocycle progression, we first determined an S phase index for each mutant genotype. Whereas approximately 60% of Histone Wild Type control (*HWT*; n = 211) and *H4K16R* mutant (n = 284) polytene chromosome spreads were in S phase (i.e. PCNA-positive) (Figure 3.1B), only ~30% of *H3K9R* (n = 286) mutant salivary glands were in S phase (Figure 3.1B; $p < 0.0001$). These data suggest that H3K9 is required for proper endocycle progression in the *Drosophila* salivary gland whereas H4K16 is dispensable. Consistent with this interpretation, *H3K9R* mutant salivary glands are smaller than *HWT* control glands at early larval stages (Figure 3.1C). However, because development is delayed by ~24 hours in *H3K9R* mutants (Penke et al. 2016), they eventually attain a similar size as *HWT* control glands by the pre-wandering stage (Figure 3.1C).

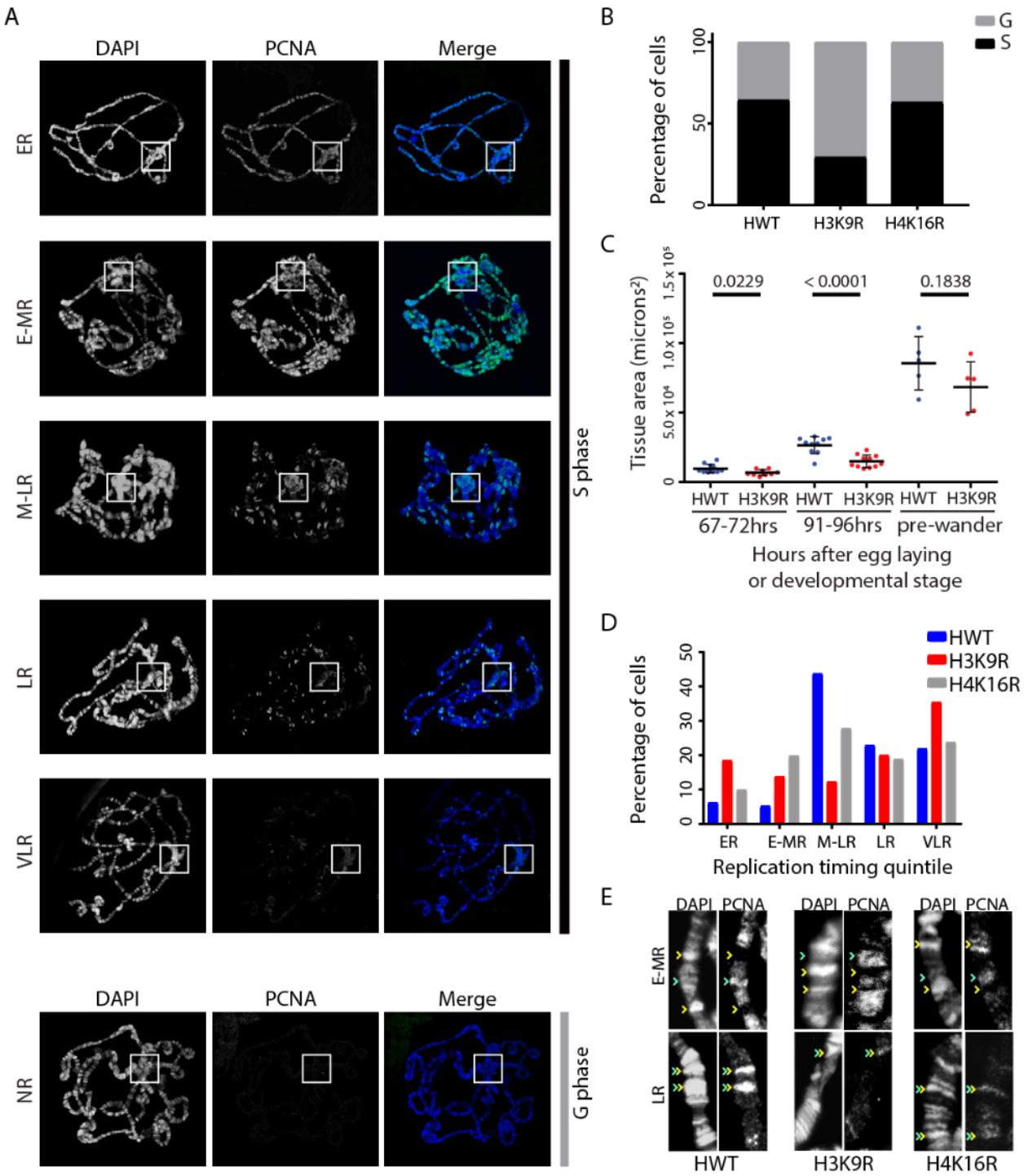
In the salivary gland polytene chromosome spreads, the largely euchromatic chromosome arms extend from a single chromocenter composed of the pericentric heterochromatin from each of the four chromosomes (Figure 3.1A). Genome-wide patterns of active replication change throughout the duration of S phase, with chromosome arms replicating earlier and the chromocenter replicating later. In addition, the euchromatin along chromosome arms (DAPI-dim inter-bands; Figure 3.1E) replicates earlier than intercalary heterochromatin (DAPI bright bands; Figure 3.1E), resulting in changing patterns of PCNA staining that can be used to monitor S phase progression. As performed previously

(Kolesnikova et al. 2013; Andreyeva et al. 2017), we binned polytene chromosome spreads from pre-wandering third-instar larvae into one of six RT categories based on the pattern of active replication as determined by PCNA staining: early-replicating (ER), early/mid-replicating (E-MR), mid/late-replicating (M-LR), late-replicating (LR), very late-replicating (VLR), and non-replicating (NR) (Figure 3.1A; Materials and Methods). If either histone mutation affected endoreplication, such as the time during S phase when either euchromatin or heterochromatin replicates (i.e. replication timing or RT), we would expect a change in PCNA staining patterns and/or a change in the distribution of these categories relative to control.

H4K16 acetylation is found in euchromatic regions of the genome while H3K9 methylation is enriched in heterochromatin (Kharchenko et al. 2011). Consequently, *H4K16R* and *H3K9R* mutants might influence earlier and later replicating regions of the genome, respectively. However, we recently demonstrated that genome-wide RT in female diploid wing imaginal disc cells is unchanged in *H4K16R* mutants (Armstrong et al. 2018). Furthermore, pericentric heterochromatin in these cells generally remains late replicating in *H3K9R* mutants, although select domains within the pericentromeres advance replication timing (Armstrong et al. 2018). Despite modest RT effect in diploid cells of these histone mutant wing discs, we next asked whether H3K9 and H4K16 directly contribute to RT during endoreplication in the salivary gland. Our single-blind analysis revealed a significant difference in RT pattern distributions between both *H3K9R* ($n = 65$; $p < 0.00001$) and *H4K16R* ($n = 101$; $p = 0.004956$) when compared to *HWT* control ($n = 96$) (Figure 3.1D). The *H4K16R* and *HWT* RT distributions were similar, with M-LR as the most prevalent RT category in each genotype. However, the *H4K16R* RT distribution shifts from M-LR to

earlier replicating patterns (ER and E-MR) relative to *HWT* control (Figure 3.1D). In contrast, the M-LR RT category was the least prevalent in *H3K9R* mutants, and the overall distribution of *H3K9R* RT categories was obviously different than either *HWT* or *H4K16R*. This distribution shows a biphasic shift, with increases in both earlier (ER and E-MR) and VLR RT patterns relative to *HWT* control (Figure 3.1D). High magnification images revealed that for both histone mutants PCNA staining in early-mid S phase chromosomes occurs in DAPI-dim inter-bands, and in late S phase PCNA staining occurs in DAPI-bright bands, as in the *HWT* control (Figure 3.1E). These data suggest that the timing of replication of intercalary heterochromatin is not advanced in either histone mutant, and that large-scale changes in RT are not solely responsible for the change in distribution of RT categories.

Figure 3.1. H3K9 promotes endoreplication of the *Drosophila* salivary gland. A) Polytene chromosome spreads from *HWT* pre-wandering third-instar larvae stained for PCNA (green) and DAPI to detect DNA (blue). Representative early-replicating (ER), early/mid-replicating (E-MR), mid/late-replicating (M-LR), late-replicating (LR), very late-replicating (VLR) and non-replicating (NR) PCNA patterns are shown. White boxes designate the chromocenter as identified by HP1a staining (*HWT* and *H4K16R*) or cytologically (*H3K9R*) (not shown). **B**) Percentage of PCNA-positive (S phase; black) and PCNA-negative (G phase; grey) polytene chromosome spreads for *HWT* (n = 211), *H3K9R* (n = 286; $p < 0.001$) and *H4K16R* (n = 284; $p > 0.05$) genotypes. Significance was determined using the Chi-squared test. All S phase measurements were taken at the pre-wandering developmental stage. **C**) Salivary gland area of *HWT* and *H3K9R* at 67–72 hours after egg deposition ($p = 0.0229$), 91–96 hours after egg deposition ($p < 0.0001$) and at the pre-wandering third-instar larval stage ($p = 0.1838$) (Student's T test). D) Percentage of PCNA-positive polytene chromosome spreads in each of the five replication timing pattern categories shown in A for *HWT* (n = 96), *H3K9R* (n = 65; $p < 0.00001$) and *H4K16R* (n = 101; $p = 0.004956$) genotypes. Significance determined using the Chi-squared test and *HWT* data as the expected categories. E) Representative chromosome arms with E-MR and LR PCNA-patterns for *HWT, H3K9R* and *H4K16R*. In E-MR patterns, PCNA colocalizes with DAPI-dim inter-bands (green arrowheads) but not DAPI-bright bands (yellow arrowheads). In LR patterns, PCNA colocalizes with DAPI-bright bands (green/yellow double arrowheads). Experiments were performed in collaboration with Taylor Penke, Samuel Chao, and Gabrielle Gentile.


A
DAPI
PCNA
Merge
ER
E-MR
M-LR
LR
VLR
S phase
NR
G phase
B
G
S
Percentage of cells
100
50
0
HWT
H3K9R
H4K16R
C
Tissue area (microns²)
0.0229
< 0.0001
0.1838
HWT H3K9R HWT H3K9R HWT H3K9R
67-72hrs
91-96hrs
pre-wander
Hours after egg laying
or developmental stage
D
HWT
H3K9R
H4K16R
Percentage of cells
ER
E-MR
M-LR
LR
VLR
Replication timing quintile
E
DAPI PCNA
E-MR
LR
HWT
H3K9R
H4K16R

We posited that the increased proportion of *H3K9R* mutant chromosomes with a VLR PCNA pattern (where replication is occurring primarily at the chromocenter) represents more extensive replication of normally under-replicated sequences at the pericentromeres. This hypothesis predicts that following completion of all endoreplication cycles, *H3K9R* mutants would have increased copy number at normally under-replicated sequences compared to both *H4K16R* mutants and *HWT* controls. To test this hypothesis, we subjected genomic DNA isolated from *H3K9R*, *H4K16R*, and *HWT* wandering third-instar larval salivary glands to Illumina sequencing (Figure 3.2A-C; Figure 3.3A-D). At this developmental stage, endoreplication cycles have ceased and salivary gland cells have reached their final ploidy. DNA copy number profiles from two biological replicate samples were generated by determining the normalized read count of paired-end 100bp reads at 10kb windows tiled across the genome. The replicate samples from each of the three genotypes correlated well with each other (Figure 3.3D). Therefore, for all subsequent analyses, we used the averaged reads per million (RPM) normalized values of the two replicates per genotype (Figure 3.3A-C).

To identify DNA copy number differences genome-wide, we determined normalized copy number values at 10kb windows along all major chromosome scaffolds (2L, 2R, 3L, 3R, 4, and X) for *H3K9R* or *H4K16R* and *HWT* control (Figure 3.3A-C). The data were plotted as a $\log_2$ transformed mutant/HWT ratio of normalized read counts, resulting in a relative copy number change for each 10kb window (Figure 3.2C). The most striking feature of these data is an increase in DNA copy number at *H3K9R* pericentromeres ($p < 2.2 \times 10^{-16}$; Student's T test) (Figure 3.2A,C; Figure 3.3A-B). Some *H3K9R* pericentric regions are enriched sixteen-fold relative to *HWT* control. These whole genome sequencing data suggest

that increased copy number at pericentromeres may contribute to the observed disorganization and enlargement of *H3K9R* chromocenters that we reported previously (Figure 3.2D) (Penke et al. 2016). In addition, when considered as a whole, the DNA copy number along chromosome arms in *H3K9R* mutants shows a slight, but statistically significant ($p < 2.2 \times 10^{-16}$), decrease relative to *HWT* control (Figure 3.2C). This decrease may result from a larger proportion of *H3K9R* mutant reads mapping to pericentric regions, causing a corresponding relative decrease in the number of reads for other genomic regions, and thus is not likely biologically meaningful. In contrast to the *H3K9R* results, we find that DNA copy number in *H4K16R* mutants is not significantly different than *HWT* controls either at pericentric regions or along chromosome arms (Figure 3.2B,C; Figure 3.3A,C). In agreement with these data, the *H4K16R* chromocenter is cytologically similar to the *HWT* control (Figure 3.2D,E).

Figure 3.2. DNA copy number in pericentric heterochromatin is elevated in *H3K9R* mutants.

A,B) Heatscatter plot of A) *H3K9R*/*HWT* $\log_2$ ratio and B) *H4K16R*/*HWT* $\log_2$ ratio of normalized copy number at 10kb windows along Chromosomes 2 and 3. LOESS regression line of modENCODE H3K9me2 ChIP signal is shown in red (GSE47260). **C**) Quantification of mutant/*HWT* ratio of normalized copy number at 10kb windows for all major chromosome scaffolds (Chromosomes 2L, 2R, 3L, 3R, 4 and X) separated into pericentromeres (Peri) and chromosome arms (Arms) (* = $p < 0.001$; Student's T test). Coordinates for pericentromeres and chromosome arms were defined in References (Riddle et al. 2011; Hoskins et al. 2015) (see also Figure 3.3). **D**) Representative polytene chromosome chromocenter from *HWT*, *H3K9R* and *H4K16R* wandering third-instar salivary glands stained with DAPI. Scale bar = 10 μM. **E**) Quantification of cytological categories for *HWT* and *H4K16R* chromocenters as performed in Reference (Penke et al. 2016). *HWT* and *H4K16R* chromocenters shown in panel D represent the organized category whereas the *H3K9R* chromocenter shown represents the severely disorganized category, which we previously reported comprises 72% of *H3K9R* chromocenters (Penke et al. 2016). The distribution of chromocenters among the three categories between *HWT* and *H4K16R* is not statistically different ($p > 0.0001$; Chi squared test). Experiments were performed in collaboration with Taylor Penke.

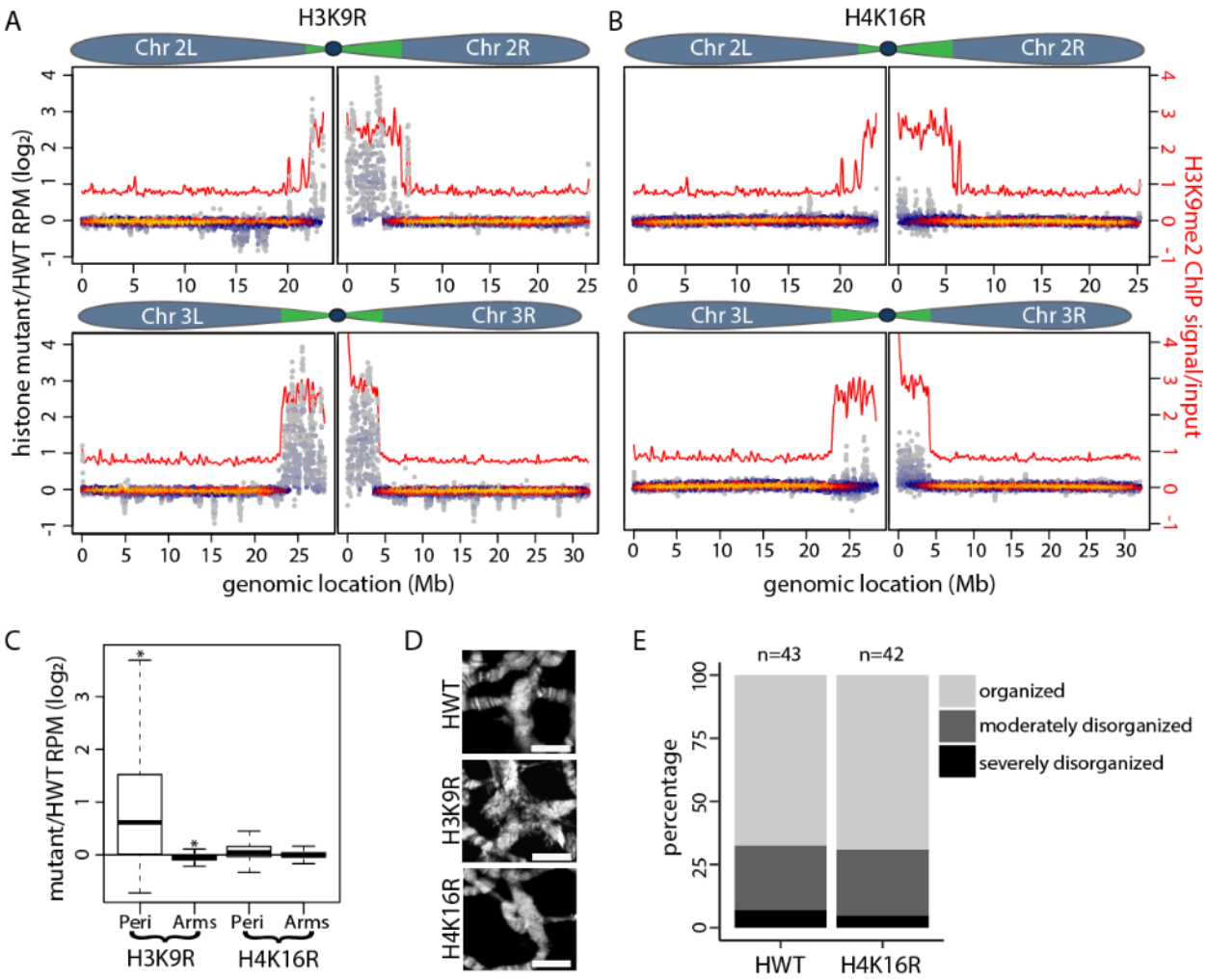
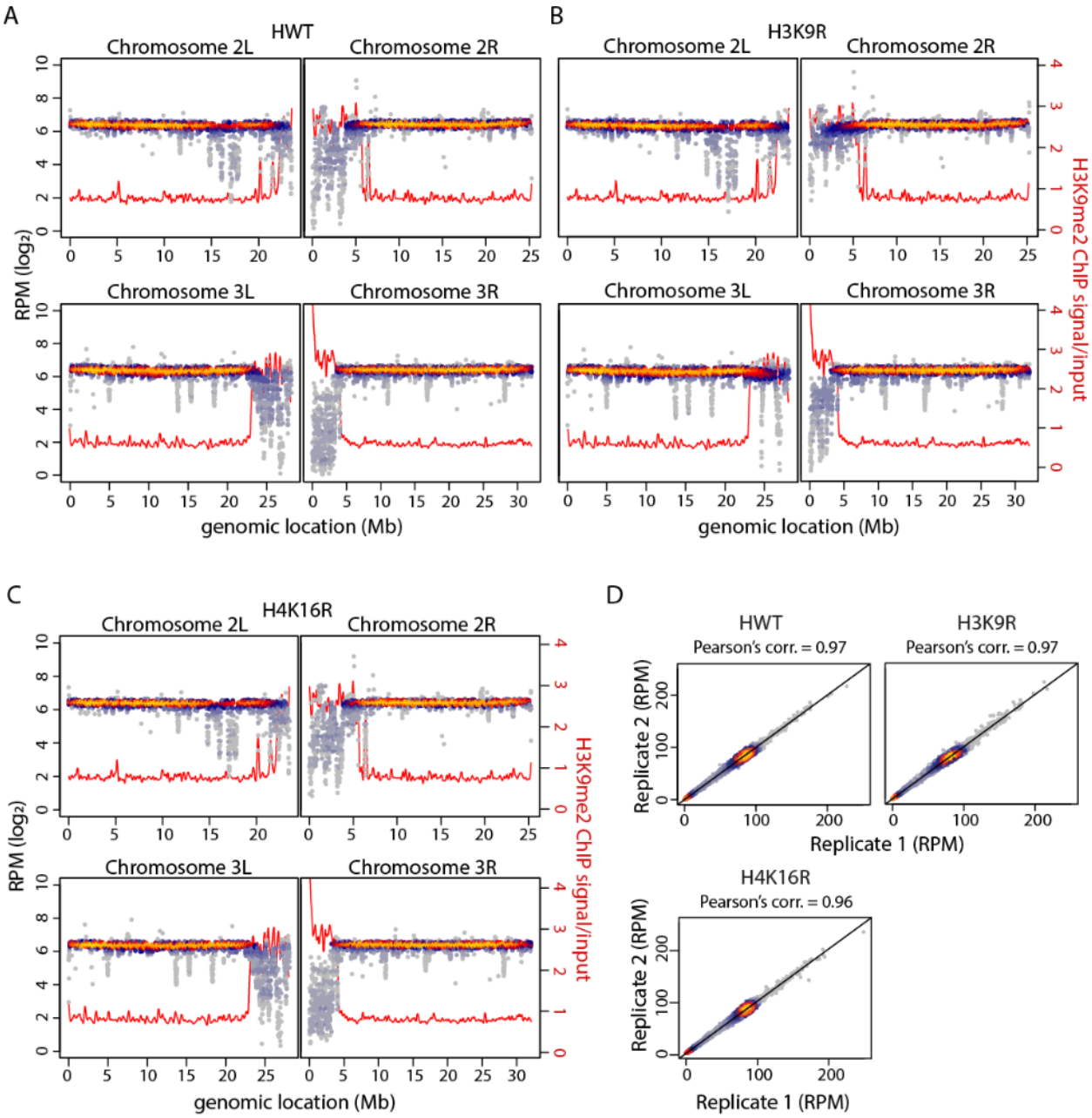

A
H3K9R
Chr 2L
Chr 2R
Chr 3L
Chr 3R
histone mutant/HWT RPM (log2)
genomic location (Mb)
B
H4K16R
H3K9me2 ChIP signal/input
C
mutant/HWT RPM (log2)
Peri
Arms
H3K9R
H4K16R
D
HWT
H3K9R
H4K16R
E
n=43
n=42
percentage
organized
moderately disorganized
severely disorganized
HWT
H4K16R

Figure 3.3. DNA copy number in pericentric heterochromatin is elevated in *H3K9R* mutants.

A–**C**) Heatscatter plot of A) *HWT* normalized copy number (log_2), **B**) *H3K9R* normalized copy number (log_2) and **C**) *H4K16R* normalized copy number (log_2) at 10kb windows along Chromosomes 2 and 3. LOESS regression line of modENCODE H3K9me2 ChIP signal is shown in red (GSE47260). **D**) Heatscatter plot comparing normalized signal at 10kb windows of two replicates each for *HWT*, *H3K9R* and *H4K16R* genotypes along all major chromosome scaffolds (Chromosomes 2L, 2R, 3L, 3R, 4 and X). Pearson's correlation value is indicated. Experiments were performed in collaboration with Taylor Penke.


A
HWT
Chromosome 2L
Chromosome 2R
Chromosome 3L
Chromosome 3R
RPM (log2)
genomic location (Mb)
B
H3K9R
H3K9me2 ChIP signal/input
C
H4K16R
D
HWT
H3K9R
Pearson's corr. = 0.97
Pearson's corr. = 0.97
Replicate 2 (RPM)
Replicate 1 (RPM)
H4K16R
Pearson's corr. = 0.96

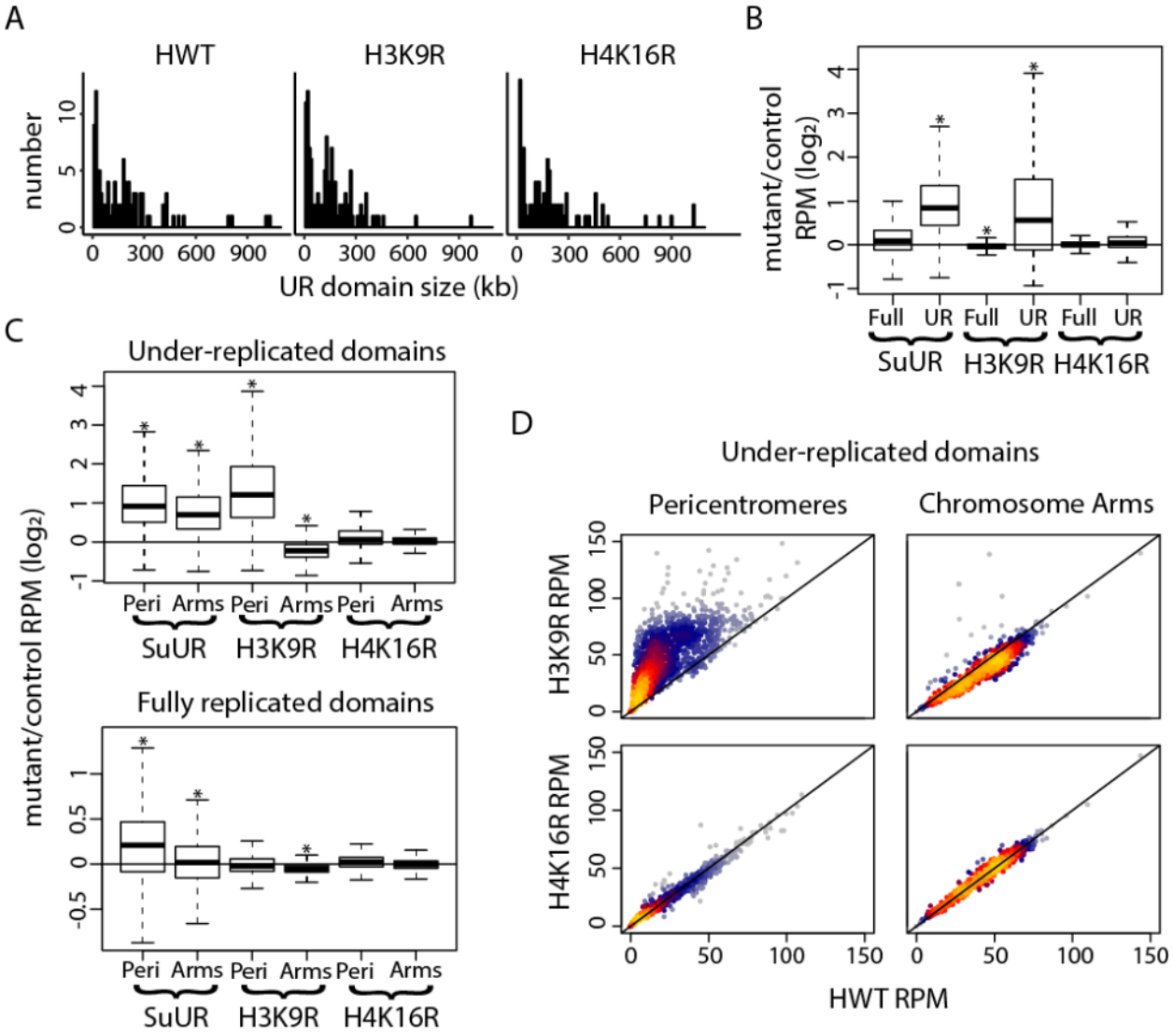
These whole genome sequencing data suggest that *H3K9R* mutants, but not *H4K16R* mutants, are defective in under-replication. To assess this possibility more directly, we next identified under-replicated domains in *H3K9R*, *H4K16R*, and *HWT* using CNVnator, a highly sensitive method for copy number variation detection based on a statistical analysis of read depth of short reads (Abyzov et al. 2011). This method was used previously in *Drosophila* (Munden et al. 2018). We required that under-replicated domains called by CNVnator be greater than 10kb. Using this criterion, we detected 101 under-replicated domains in *HWT* control salivary glands, 86 and 98 of which overlap with domains identified in *H3K9R* and *H4K16R* mutants, respectively. The size distribution of under-replicated domains in each of the three genotypes were similar, with sizes ranging from 11kb to over a megabase and medians ranging from 127kb to 138kb (Figure 3.4A).

We used the 101 under-replicated domains identified in *HWT* salivary glands to more closely investigate under-replication in *H3K9R* and *H4K16R* mutants. We also included in our analysis previously published Illumina sequencing data from *SuUR* mutant salivary glands, which have reduced under-replication both at pericentromeres and along chromosome arms (Nordman et al. 2014). When we use 10kb windows to compare DNA copy number at under-replicated domains between each mutant genotype and their respective controls, we observe a significant reduction of under-replication in both *SuUR* and *H3K9R* mutants ($p < 2.2 \times 10^{-16}$; Student's T test) (Figure 3.4B). In addition, we found a small, but statistically significant, decrease in copy number at fully replicated regions in *H3K9R* mutants ($p = 8.25 \times 10^{-7}$; Student's T test), which as noted above most likely results from the way the sequencing data was analyzed rather than a biological phenomenon (Figure 3.4B). In contrast, we found no significant changes in copy number at either subset of genomic regions

in *H4K16R* mutants (Figure 3.4B). These data suggest that copy number differences between *H3K9R* and *HWT* are due to failure of the normal under-replication mechanism.

We next partitioned the 101 under-replicated regions identified in *HWT* into those located on chromosome arms and those located within pericentric heterochromatin. For this purpose, pericentric heterochromatin was defined by high levels of H3K9me2 enrichment as described previously (Riddle et al. 2011; Hoskins et al. 2015). When considering all under-replicated domains within pericentric heterochromatin, we observed a significant increase in copy number in *H3K9R* mutants ($p < 2.2 \times 10^{-16}$) relative to *HWT* control (Figure 3.4C,D). Furthermore, 10/48 (21%) under-replicated domains identified in *HWT* pericentric heterochromatin were not identified using CNVnator in *H3K9R* mutants, suggesting a strong suppression of under-replication in these 10 domains. When considering copy number at under-replicated domains along chromosome arms, we observed a small but statistically significant ($p < 4.429 \times 10^{-10}$) decrease in copy number in *H3K9R* mutants relative to *HWT* control (Figure 3.4C,D). This decrease in copy number may result from the 18 newly identified under-replicated domains in *H3K9R* that were not identified in *HWT* or may result from the sequencing analysis as noted above. In contrast to these data, *SuUR* mutants have reduced under-replication both at pericentric heterochromatin and along chromosome arms (Figure 3.4C) (Nordman et al. 2014; Munden et al. 2018). We found no significant changes in copy number either at under-replicated regions in pericentric heterochromatin or along chromosome arms in *H4K16R* mutants (Figure 3.4C,D). These data demonstrate that H3K9 is required for normal under-replication at pericentric heterochromatin and suggest that under-replication along chromosome arms occurs in an H3K9-independent manner.

Figure 3.4. Under-replication of pericentric heterochromatin is H3K9-dependent. **A**) Histogram of under-replicated domains identified by CNVnator (Abyzov et al. 2011) for *HWT*, *H3K9R* and *H4K16R* genotypes. Bin size is set to 10kb. **B**) Boxplot of the *SuUR/OregonR*, *H3K9R*/*HWT* and *H4K16R*/*HWT* $\log_2$ ratios of DNA copy number at 10kb windows at fully replicated (Full) and under-replicated (UR) domains as defined by CNVnator (Abyzov et al. 2011) in *HWT* (* = $p < 0.001$; Student's T test). **C**) Boxplot of the *SuUR/OregonR*, *H3K9R*/*HWT* and *H4K16R*/*HWT* $\log_2$ ratios of normalized signal at 10kb windows at under-replicated domains (top panel) and fully replicated domains (bottom panel) in pericentromeres (Peri) or chromosome arms (Arms) (* = $p < 0.001$; Student's T test). **D**) Heatscatter plot of normalized signal at 10kb windows at under-replicated domains in *HWT* versus *H3K9R* (top panels) or versus *H4K16R* (bottom panels) at pericentromeres (left panels) and chromosome arms (right panels). Experiments were performed in collaboration with Taylor Penke.


A
HWT
H3K9R
H4K16R
number
UR domain size (kb)
B
mutant/control RPM (log2)
Full UR Full UR Full UR
SuUR H3K9R H4K16R
C
Under-replicated domains
mutant/control RPM (log2)
Peri Arms Peri Arms Peri Arms
SuUR H3K9R H4K16R
Fully replicated domains
D
Under-replicated domains
Pericentromeres
Chromosome Arms
H3K9R RPM
H4K16R RPM
HWT RPM

Discussion

Here, we utilized a genetic platform for histone gene replacement to interrogate the function of replication-dependent H3K9 and H4K16 in *Drosophila* salivary gland endoreplication. We found that while H3K9 is important for salivary gland endoreplication, H4K16 is largely dispensable. We observed three phenotypes in *H3K9R* mutant salivary glands: i) a decrease in the S phase index, ii) a biphasic shift in replication timing toward both earlier and the very latest (i.e. chromocenter replication) patterns, and iii) a reduction in the level of under-replication at pericentric heterochromatin, but not along chromosome arms.

A decrease in the S phase index could result from a reduction in the number or duration of endo-S phases, an increase in the duration of G phase, or both. The *H3K9R* mutant salivary glands reach the same size as control, consistent with completion of the endoreplication program. In addition, the slower development of *H3K9R* mutant animals may result in longer G periods between endo S phases, thus further reducing the S phase index. Another possibility is that the endoreplication timing program is generally condensed in *H3K9R* mutants such that S phase occurs more quickly, accounting for the change in distribution of RT categories and contributing to the decrease in S phase index. These changes unlikely indirectly result from changes in transcription, as we did not detect significant changes in the expression of protein coding genes, including those encoding replication factors, in the replication-dependent *H3K9R* mutant (Penke et al. 2016; Armstrong et al. 2018).

The latest replicating sequences of polyploid salivary gland cells are not fully replicated each endocycle, yielding regions of the genome with decreased copy number,

particularly in pericentric heterochromatin but also at specific loci along chromosome arms (Spradling and Orr-Weaver 1987). We suggest that the elevated number of *H3K9R* mutant salivary gland nuclei with a very late replication pattern, represented by PCNA staining of the chromocenter, results from more extensive replication of pericentric heterochromatin each endocycle. Consistent with such a failure of normal under-replication, in *H3K9R* mutants we detected an increase in DNA copy number at pericentric heterochromatin by whole genome sequencing, as well as altered chromocenter cytology (Penke et al. 2016). Interestingly, we did not detect decreased under-replication along chromosome arms, indicating that replication-dependent H3K9 is particularly important for endoreplication control in pericentric heterochromatin. This result is reminiscent of our previous observations that replication-dependent H3K9R mutation disrupts HP1a recruitment, nucleosome occupancy, and transposon repression at pericentric heterochromatin in diploid wing imaginal discs, without appreciably affecting the function of euchromatin (Penke et al. 2016; Armstrong et al. 2018; Penke et al. 2018). An important caveat to our observations is that we cannot rule out a contribution from K9 of the variant histone H3.3 to under-replication along salivary gland polytene chromosome arms. Animals in which both H3.3 and replication-dependent H3 contain K9R mutations die as early first instar larvae, precluding the appropriate genetic experiment (Penke et al. 2018). Similarly, our *H4K16R* analyses cannot rule out a small contribution from the replication-independent *His4r* gene, which resides outside the replication-dependent histone gene cluster and encodes a protein identical to replication-dependent H4 (Armstrong et al. 2018; Copur et al. 2018).

Our previous analysis of *H3K9R* diploid wing discs revealed only a small number of advanced RT changes within pericentric heterochromatin (Armstrong et al. 2018). In

contrast, our analyses here revealed that most under-replicated domains within pericentric heterochromatin increased in DNA copy number in *H3K9R* mutant polyploid salivary glands. These data suggest that replication-dependent H3K9 plays a more significant role in regulating under-replication during salivary gland endoreplication than in regulating late replication during the canonical diploid cell cycle. Alternatively, these processes might be controlled by distinct mechanisms. The biological function of under-replication is not known, and thus the consequence of losing pericentric under-replication is uncertain. Interestingly, *Rif1* mutants, which lack under-replication altogether (Munden et al. 2018), are viable and fertile (Munden et al. 2018; Seller and O'Farrell 2018).

What might be the mechanism by which H3K9R promotes under-replication of pericentric heterochromatin? Previous studies established SuUR as a key regulator of under-replication in polyploid genomes of *Drosophila* salivary glands (Belyaeva et al. 1998; Makunin et al. 2002; Nordman et al. 2014; Pokholkova et al. 2015; Munden et al. 2018). Unlike *H3K9R* mutants, under-replication in *SuUR* mutants is reduced both along chromosome arms and at pericentric heterochromatin (Makunin et al. 2002). The mode of SuUR association with these two regions of the genome is different, being SNF2 domain-dependent and dynamic with replication forks and SNF2 domain-independent and more constitutive within pericentric heterochromatin (Munden et al. 2018). SuUR forms a protein complex with Rif1, which recruits Protein Phosphatase 1 (PP1), and these interactions are required to promote under-replication (Munden et al. 2018). In addition, HP1a and SuUR depend on one another for chromatin association (Pindyurin et al. 2008; Pokholkova et al. 2015). Thus, one potential explanation for reduced pericentric heterochromatin under-replication in *H3K9R* mutants is the loss of pericentric HP1a (Penke et al. 2016). In the

absence of HP1a, SuUR's constitutive association with pericentric heterochromatin is reduced (Pindyurin et al. 2008), which may prevent the downstream effectors of under-replication, Rif1 and PP1, from properly suppressing replication at these regions of the genome.

In conclusion, our data indicate that under-replication at salivary gland pericentric heterochromatin occurs through an H3K9-dependent mechanism, and therefore suggest that the hallmarks of constitutive heterochromatin, H3K9me and HP1a, are critical regulators of under-replication in *Drosophila* polyploid cells.

Acknowledgements

This work was supported by NIH Grants R01-GM124201 to R.J.D.; F31-GM115194 to T.J.R.P; R35-GM128851 to D.J.M.; and R01-GM129132 to A.G.M. In addition, R.L.A. and T.J.R.P. were supported in part by an NIH predoctoral training grant T32-GM007092. We thank the High Throughput Sequencing Core Facility, supported in part by P30 CA016086 Cancer Center Core Support Grant to the UNC Lineberger Comprehensive Cancer Center. We thank Alex Munden and Jared Nordman (Vanderbilt University) for bioinformatics analysis advice.

CHAPTER 4- RIF1 FUNCTIONS IN A TISSUE-SPECIFIC MANNER TO CONTROL REPLICATION TIMING THROUGH ITS PP1-BINDING MOTIF

Introduction

DNA replication initiates from discrete regions of the eukaryotic genome, known as replication domains, in a precise chronological manner during S phase. This temporal order of DNA replication is known as the DNA replication timing (RT) program and is evolutionarily conserved from yeast to humans (Rivera-Mulia and Gilbert 2016). In metazoan species, replication domain sizes range from hundreds of bases to megabases, and their RT is correlated with transcriptional activity, chromatin structure, and position within the nucleus (MacAlpine et al. 2004; Schwaiger et al. 2009; Eaton et al. 2011; Rivera-Mulia and Gilbert 2016; Almeida et al. 2018; Heinz et al. 2018). Furthermore, RT domains are highly correlated with topologically associated domains (TADs), where a near one-to-one correlation has been observed between RT domains and TADs (Pope et al. 2014). While RT is clearly influenced by chromatin structure and nuclear organization, the exact function of RT is not fully understood. Importantly, defects in RT are associated with genome instability, and RT is often altered in cancer cells (Stamatoyannopoulos et al. 2009; Koren et al. 2012; Donley and Thayer 2013). Therefore, understanding the processes and factors that contribute to RT is key to understanding fundamental aspects of eukaryotic DNA replication and genome stability.

Both cellular differentiation and cellular identity influence genome-wide RT, suggesting that the underlying mechanisms regulating RT are plastic during development.

Comparison of genome-wide RT between three lines of cultured *Drosophila* cells revealed differences in RT across ~8% of the genome (Lubelsky et al. 2014). More extensive RT profiling using *in vitro* models of cellular differentiation from multiple mammalian cell lineages has revealed ~50% of the genome is subject to cell-type specific RT changes (Hiratani et al. 2008; Hiratani et al. 2010). Furthermore, in mammalian cells, the RT program goes through a global reorganization where many small RT domains consolidate into larger RT domains as cells differentiate from embryonic stem cells to more differentiated cell types (Ryba et al. 2010). It is still unclear, however, whether cell-type specific changes in RT are developmentally programmed directly or whether differential RT is a passive reflection of the changes in chromatin structure and nuclear organization that occur during cellular differentiation.

Multiple *trans*-acting replication factors control RT from yeast to humans. Loading of the MCM replicative helicase during G1 phase of the cell division cycle and helicase activation during S phase are key steps in RT control (Bell and Stillman 1992; MacAlpine et al. 2010; Mantiero et al. 2011; Collart et al. 2013; Miotto et al. 2016). Several factors are limiting for replication initiation (Sld2, Sld3, Dpb11, Dbf4 and Cdc45) and their overexpression disrupts RT in budding yeast and *Xenopus* (Mantiero et al. 2011; Collart et al. 2013). A critical *trans*-acting RT-regulating factor is Rif1 (Rap1-interacting factor 1), which controls RT from yeasts to humans (Cornacchia et al. 2012; Hayano et al. 2012; Yamazaki et al. 2012; Peace et al. 2014; Foti et al. 2016). In animals, it is not clear whether the genomic regions that Rif1 targets during differentiation are cell-type specific or whether Rif1 selectively regulates specific regions of the genome regardless of cell type. Although Rif1 is only modestly conserved, all Rif1 orthologs contain a Protein Phosphatase 1 (PP1)-

interaction motif, suggesting that PP1 recruitment is a critical function of Rif1. Rif1 recruitment of PP1 to chromatin prevents the Dbf4-dependent kinase (DDK) activation of DNA-loaded helicases (Davé et al. 2014; Hiraga et al. 2014; Mattarocci et al. 2014; Hiraga et al. 2017; Sukackaite et al. 2017). However, how specific loss of the Rif1-PP1 interaction affects RT genome wide has not been determined.

To better understand the extent to which Rif1 regulates RT in various unperturbed cell types during development, we have measured RT in the *Drosophila* larval wing discs and adult ovarian follicle cells in the presence and absence of Rif1. Here, we identify regions of the genome that change RT as a function of cell lineage and determine Rif1-dependent changes in RT in different tissue types. We found that cell lineage is a major driver of RT and demonstrate that tissue-specific transcription is not a major contributor to tissue-specific RT. Importantly, although RT in a subset of the genome depends on Rif1 similarly in different tissues, Rif1 largely acts in a tissue-specific manner to control RT. Additionally, the Rif1-PP1 interaction motif is required for Rif1-dependent control of RT, suggesting that PP1 recruitment to replicative helicases is the predominant mechanism Rif1 utilizes for RT control.

Materials and Methods

FACS and genomic DNA sequencing

Isolated nuclei from *OregonR*, $Rif1^{1}/Rif1$ *(Rif1*$^{-}$*)*, and $Rif1^{PP1}/Rif1$ *(*$Rif1^{PP1}$*)* female adult ovaries and *yw*, *Rif1*$^{-}$, and $Rif1^{PP1}$ female 3rd instar larval wing imaginal discs from were sorted into G1 and S populations by a FACSAria II or III based on DAPI intensity and subsequently pelleted, flash frozen, and stored at -80°C prior to DNA isolation and library

preparation. Libraries were prepared with the Rubicon ThruPLEX DNA-seq kit for wing imaginal disc samples and with the NEBNext Ultra II DNA Library Prep kit for follicle cell samples and subjected to Illumina HiSeq 2500 single-end 50bp sequencing for wing imaginal disc samples and Illumina HiSeq X paired-end 150bp sequencing for follicle cell samples.

RT Characterization

Reads from G1 and S samples were aligned to the dm6 reference genome (Release 6.04) using Bowtie 2 (v2.3.2) default parameters (Langmead et al. 2009). Reads with a MAPQ score greater than 10 were retained using SAMtools (v1.9) (Li et al. 2009). BEDTools coverage (v2.26.0) was used to quantify the number of reads mapping to each 100kb window, with results normalized to read depth (Quinlan and Hall 2010). Replication timing (RT) values were obtained by averaging the S/G1 ratio of reads per million (RPM) value from each S phase replicate for a particular window size. Profiles were generated by plotting the RT value at each window versus genomic location. Quantile normalization was performed for comparisons between samples through the preprocess Core R package to equalize the dynamic range of RT values (Bolstad 2016). The limma statistical package was used to identify 100kb windows with significantly altered RT values (lmFit, p value adjusted for multiple testing ($p<0.01$); absolute $\log_2$ fold change > 0.1) (Newville et al. 2014). BEDTools intersect (v2.26.0) was used to determine overlap of 100kb windows with -f 0.5 and -u parameters (Quinlan and Hall 2010). RT values and limma-generated adjusted p values at 100kb windows were used to determine median RT values and adjusted p values at 10kb windows (BEDTools map v2.26.0), and the significance threshold was adjusted at 10kb windows (p value adjusted for multiple testing ($p<0.05$); absolute $\log_2$ fold change > 0.1) (Quinlan and Hall 2010). Coordinates of chromatin states were obtained from (Kharchenko

et al. 2011) and converted to dm6 coordinates using the UCSC liftOver tool (Karolchik et al. 2004). To calculate RT domain sizes, we identified the genomic coordinates halfway between each peak and valley of an RT profile and determined the distance from one halfway point to the next.

For false discovery rate (FDR) calculations, spike-in RT bed files with 3 x 10^7 reads were generated by combining either 3 x 10^5 (1% impure), 1.5 x 10^6 (5% impure), 3 x 10^6 (10% impure), 7.5 x 10^6 (25% impure), or 1.5 x 10^7 (50% impure) randomly selected reads from each wing disc S phase replicate with 2.97 x 10^7 (1% impure), 2.85 x 10^7 (5% impure), 2.7 x 10^7 (10% impure), 2.25 x 10^7 (25% impure), or 1.5 x 10^7 (50% impure) randomly selected reads from each mitotically cycling follicle cell S phase replicate. RT profiles generated from each test dataset (1% impure, 5% impure, 10% impure, 25% impure, and 50% impure) were directly compared to RT profiles from wing discs, and differential replication timing was identified as before using the limma statistical package (lmFit, p value adjusted for multiple testing ($p<0.01$); absolute $\log_2$ fold change > 0.1) (Newville et al. 2014). We estimate that 50% of the "mitotic" follicle cell population consists of endocycling follicle cells due to the following rationale: Because the total number of follicle cells in an egg chamber after the completion of the mitotic cell divisions is 1,024, the 2C-4C population used for sorting contains 2^{10} (1,024) mitotically cycling follicle cells from all egg chambers prior to Stage 7 per ovariole and (at most) 1,024 endocycling follicle cells from the Stage 7 egg chamber per ovariole.

RNA Analyses

Follicle cell isolation, RNA extraction and sequencing: Follicle cells were isolated by trypsinizing ovaries from *OregonR* or *Rif1*^{1}*/Rif1*^{2} females as described in (Cayirliogu et al,

MCB 2003). Follicle cells were FACS sorted into TRIzol LS (Invitrogen) based on their ploidy and RNA was extracted according to the manufacture's recommendation. 250,000 – 500,000 follicle cells were used per replicate. rRNA was depleted using the RiboMinus™ Eukaryote Kit for RNA-Seq (Invitrogen) and libraries were prepared using the NEBNext® Ultra™ II RNA Library Prep.

Wing disc isolation, RNA extraction and sequencing: Total RNA was isolated from 40 *yw* and *Rif1*^{1}*/Rif1* female 3rd instar wing imaginal discs. Wing imaginal discs were homogenized in Trizol (Invitrogen) and flash frozen in liquid nitrogen. RNA was isolated using the Direct-zol RNA miniprep kit (Zymo Research). rRNA was depleted and libraries were prepared using the Ovation *Drosophila* RNA-Seq system (NuGEN). RNA isolated from *yw* wing imaginal discs was also made into libraries and sequenced with follicle cell RNA for all comparisons in Figure 4.5.

RNA seq analysis: TopHat default parameters (v2.1.1) (Trapnell et al. 2012) were used to align paired-end reads to the dm6 version of the *Drosophila* genome. Transcriptomes were generated using Cufflinks (v2.2.1, see supplementary materials for parameters). Differentially expressed transcripts were determined via edgeR statistical analysis (p value <0.01) (Robinson et al. 2010; McCarthy et al. 2012). For analyses comparison transcription to RT at 10kb windows, we either assigned the average RNA $\log_2$ fold change and average adjusted p-value from all transcripts overlapping each 10kb window or we assigned the $\log_2$ fold-change of the transcript with the lowest edgeR-generated p value at each 10kb window for analyses directly comparing RT and transcription. Results were similar irrespective of how transcription was assigned to RT windows.

Results

Cell lineage is a major driver of DNA replication timing

To analyze RT in unperturbed cell types and tissues without the need to immortalize or transform cells, we exploited the well-characterized developmental systems of *Drosophila melanogaster*. To determine how cell lineage affects RT, we generated genome-wide RT profiles from cells of two distinct *D. melanogaster* epithelial tissues: third-instar larval wing imaginal disc cells and follicle cells from female adult ovaries. Cells of the wing disc are derived from the embryonic mesoderm while ovarian follicle cells are derived from the embryonic ectoderm. To generate RT profiles, we used fluorescence-activated cell sorting (FACS) to isolate and subsequently sequence the genomes of S phase nuclei from each tissue and compared these data to those obtained from G1 phase nuclei from wing discs (Figure 4.1A; (Armstrong et al. 2018)). The premise of this method is that early-replicating DNA sequences are over-represented relative to late-replicating sequences within the S phase population. Therefore, replication timing values can be quantified by determining $\log_2$ transformed S/G1 read counts across the genome, where larger values indicate earlier replication and smaller values indicate later replication (Figure 4.1A).

To determine how lineage contributes to RT, we generated RT values at 100kb windows tiled at 10kb intervals across the genome for both wing discs and follicle cells and used a stringent significance threshold to identify differential RT between each tissue (Materials and Methods; (Armstrong et al. 2018)). RT profiles generated from individual replicates of wild type wing discs and follicle cells were strongly correlated (Pearson's correlations = 0.95 and 0.95, respectively; Figure 4.2A), whereas RT values between the two lineages were significantly more divergent (Pearson's correlation = 0.39; Figure 4.1B). While

~70% of the genome has similar RT between the two tissues, ~29% of the genome displays tissue-specific RT where 14.6% of windows replicate earlier in follicle cells and 14.5% of windows replicate earlier in wing discs (Figure 4.1C,D; Figure 4.2B; Table 4.1). Gene ontology analysis of genes located within tissue-specific RT domains did not reveal a significant enrichment of genes associated with a specific biological process. Furthermore, differential RT between wing discs and follicle cells did not preferentially affect any one chromatin state (Kharchenko et al. 2011), and replication domain sizes were highly similar between the two tissues (Figure 4.2C,D). These data demonstrate that cell lineage is a key contributor to replication timing control in *Drosophila* similar to what has been previously observed in mammalian cell culture systems (Hiratani et al. 2008; Ryba et al. 2010; Rivera-Mulia et al. 2015).

Figure 4.1. Cell lineage is a major driver of DNA replication timing in *Drosophila*. A) Experimental outline: (1) Nuclei were FACS sorted into G1 (yellow) and S (blue or green) populations based on DNA content. (2) DNA was sequenced and mapped back to the dm6 reference genome. More reads map to early than late replicating sequences. (3) S/G1 $\log_2$ ratio of mapped reads generates replication timing profiles. **B)** Heatscatter plot of wild type wing disc and wild type follicle cell S/G1 ($\log_2$) ratios at all 100kb windows using a 10kb slide across the genome. **C)** Pie chart of all 100kb windows of significantly earlier RT in wild type wing discs (green), significantly earlier RT in wild type follicle cells (blue), and unchanged RT (grey) across the major chromosome scaffolds. **D)** LOESS regression lines showing average wild type wing disc (green) and wild type follicle cell (blue) S/G1 ($\log_2$) replication timing values across the chromosome 3R scaffold. See Figure 4.2 for all other chromosome arms. Experiments were performed in collaboration with Souradip Das.

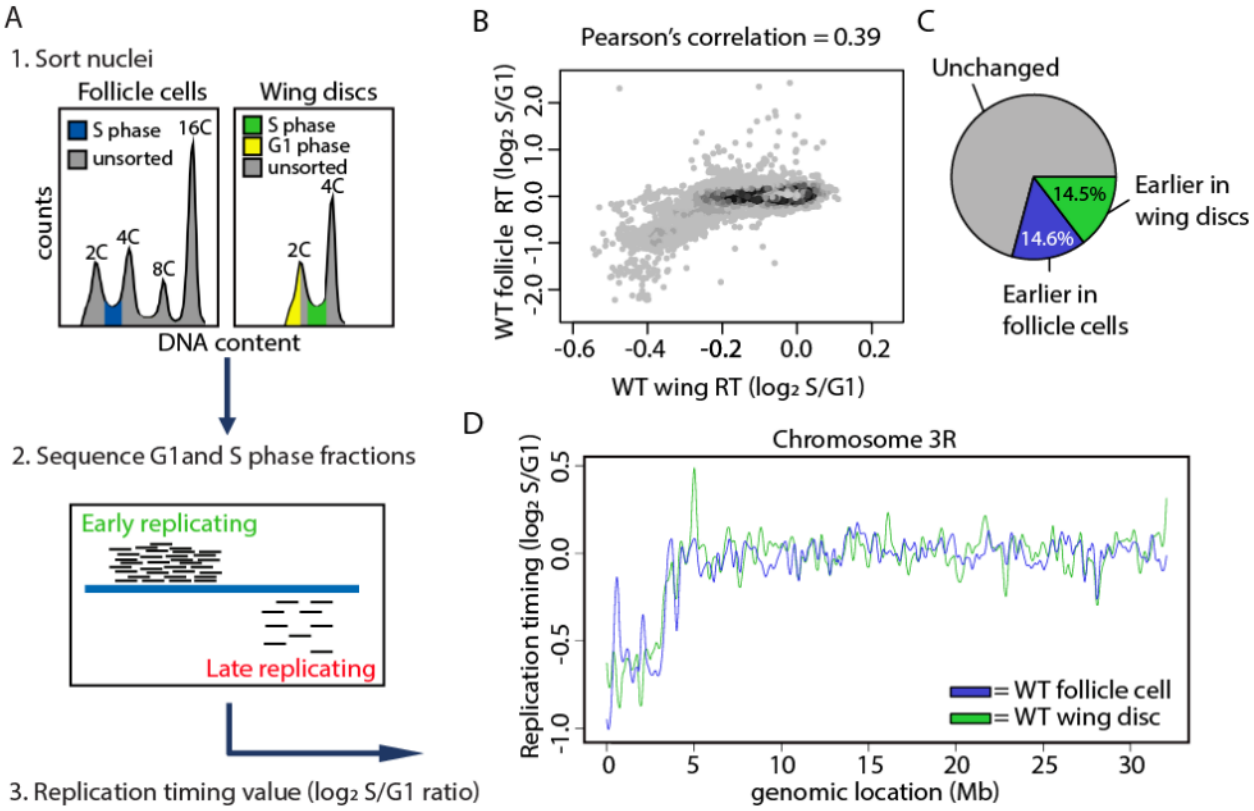

A
1. Sort nuclei
Follicle cells
Wing discs
S phase
unsorted
16C
4C
2C
8C
S phase
G1 phase
unsorted
4C
2C
counts
DNA content
2. Sequence G1and S phase fractions
Early replicating
Late replicating
3. Replication timing value (log2 S/G1 ratio)
B
Pearson's correlation = 0.39
WT follicle RT (log2 S/G1)
WT wing RT (log2 S/G1)
C
Unchanged
14.5%
Earlier in wing discs
14.6%
Earlier in follicle cells
D
Chromosome 3R
Replication timing (log2 S/G1)
genomic location (Mb)
= WT follicle cell
= WT wing disc

Figure 4.2. Characterization of RT in wild type wing discs and follicle cells. A) Heatscatter plot comparing wild type wing disc S/G1 (log_2) replicate replication timing values (top) and wild type follicle cell S/G1 (log_2) replicate replication timing values (bottom). **B)** LOESS regression line showing average S/G1 (log_2) replication timing values for wild type wings discs (green) and wild type follicle cells (blue) at 100kb windows using a 10kb slide across the major chromosome scaffolds. **C)** All 10kb windows of differential RT between wild type follicle cells and wild type wing discs were assigned to the nine chromatin states previously defined in *Drosophila* (Kharchenko et al. 2011). Shown are the percentage of each chromatin state with differential RT. **D)** Histogram of replication domain sizes in wild type follicle cells and wild type wing discs. Experiments were performed in collaboration with Souradip Das.

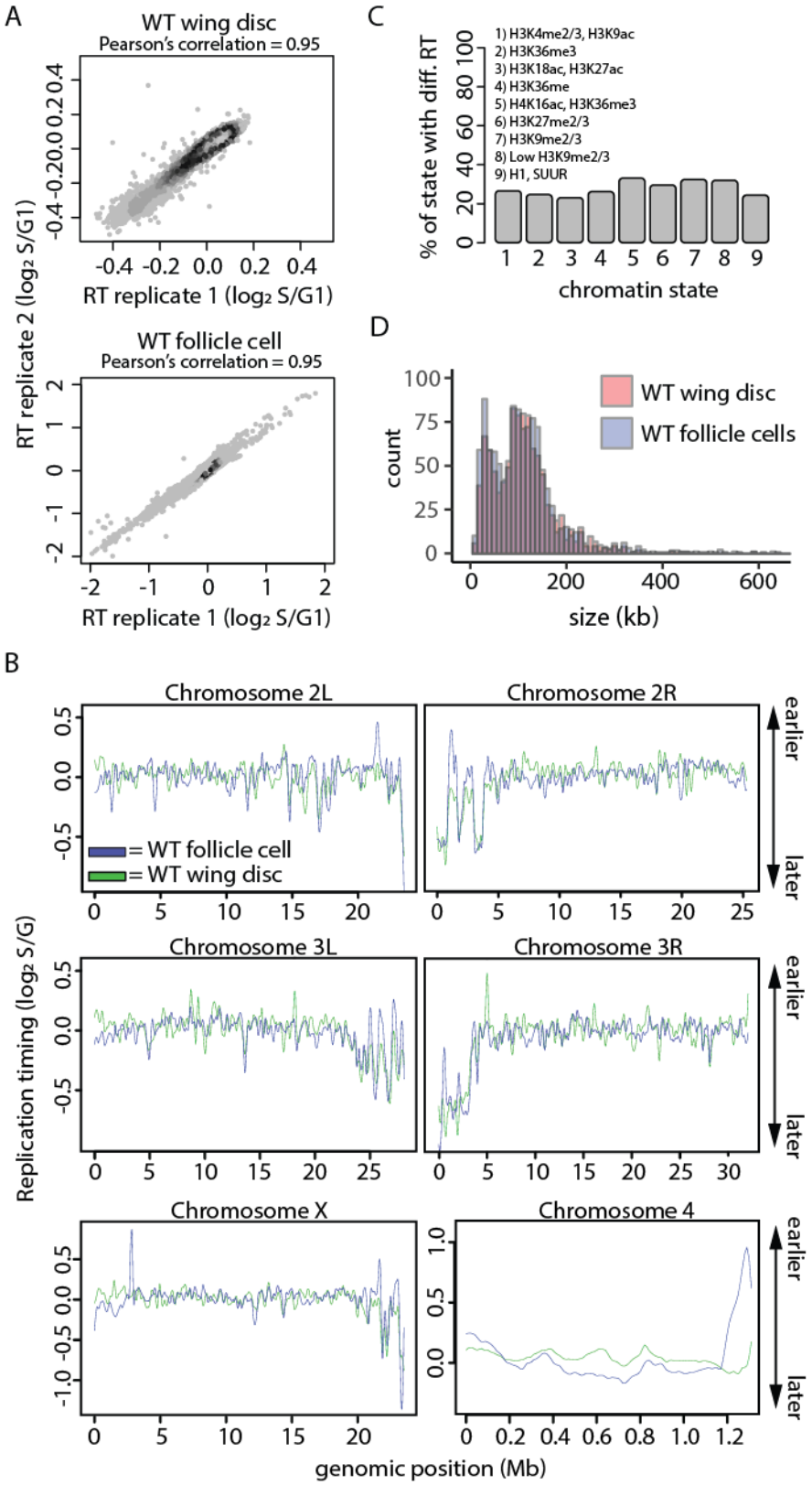

A
WT wing disc
Pearson's correlation = 0.95
RT replicate 2 (log2 S/G1)
RT replicate 1 (log2 S/G1)
WT follicle cell
Pearson's correlation = 0.95
RT replicate 2 (log2 S/G1)
RT replicate 1 (log2 S/G1)
C
1) H3K4me2/3, H3K9ac
2) H3K36me3
3) H3K18ac, H3K27ac
4) H3K36me
5) H4K16ac, H3K36me3
6) H3K27me2/3
7) H3K9me2/3
8) Low H3K9me2/3
9) H1, SUUR
% of state with diff. RT
chromatin state
D
WT wing disc
WT follicle cells
count
size (kb)
B
Chromosome 2L
Chromosome 2R
= WT follicle cell
= WT wing disc
earlier
later
Chromosome 3L
Chromosome 3R
Replication timing (log2 S/G)
Chromosome X
Chromosome 4
genomic position (Mb)

Table 4.1. Quantification of differential RT at 100kb windows using a 10kb slide across the major chromosome scaffolds.

Experiments were performed in collaboration with Souradip Das.

genotype comparison	total # of windows in comparison	# of windows advanced	# of windows delayed
WT mitotic follicle vs. WT wing disc	13391	1951 (earlier in follicle)	1943 (earlier in wing)
WT endo follilcle vs. WT mitotic follicle	13391	0	0
Rif1^{-} wing disc vs. WT wing disc	13391	552	527
Rif1^{-} mitotic follicle vs. WT mitotic follicle	13391	1095	672
Rif1$^{-/+}$ mitotic follicle vs. WT mitotic follicle	13391	478	212
Rif1^{-} endo follicle vs. WT endo follicle	13391	960	1024
Rif1^{PP1} wing disc vs. WT wing disc	13391	1726	666
Rif1^{PP1} mitotic follicle vs. WT mitotic follicle	13391	968	520

Cell type-specific transcription does not drive changes in RT

Transcriptional activity is highly correlated with RT, with early replicating regions of the genome associated with active transcription and late replicating regions associated with transcriptional repression (MacAlpine et al. 2004; Liu et al. 2012; Lubelsky et al. 2014; Rivera-Mulia and Gilbert 2016). Therefore, we determined if differences in transcriptional activity correlated with differential RT. We generated transcriptomes from wild type wing disc cells and follicle cells by total RNA-seq and identified differentially expressed transcripts between each tissue type. Individual biological replicates were highly correlated (Figure 4.3; Pearson's correlation coefficients > 0.95) and we were able to identify tissue-specific gene expression including *wingless* (*wg*) expression in wing discs and *chorion protein* (*cp*) expression in follicle cells (Figure 4.4A). We observed 3,994 differentially expressed transcripts ($p < 0.01$; edgeR) between the two tissues (Figure 4.5A), with elevated expression of 2,651 transcripts in wing discs and 1,343 transcripts in follicle cells (Figure 4.5A).

To identify whether tissue-specific RT is driven by tissue-specific gene expression between wing discs and follicle cells, we directly compared differences in RT and gene expression at 10kb windows across the genome between the two tissues. First, we compared the average change in abundance of all transcripts within each window to the RT change of that window (Materials and Methods). Although transcript abundance was modestly elevated in wing discs versus follicle cells at windows of earlier RT in wing discs (average $\log_2$ fold change = 1.45CPM), we did not observe a strong correlation between elevated gene expression and earlier RT in follicle cells (Figure 4.5B,C; Figure 4.4B). These results were consistent whether we considered 1) the average change in the abundance of all transcripts

overlapping each 10kb window (Figure 4.5B,C; Figure 4.4B), 2) the change of the most confident transcript (lowest p value) assigned to each window (Figure 4.4C), or 3) the change of the transcript with the greatest differential expression (absolute maximum $\log_2$ fold-change) assigned to each window (Figure 4.4D). Furthermore, 47.4% (791/1670) and 73.4% (813/1107) of windows with earlier RT in wing discs or follicle cells, respectively, do not contain a transcript with a significant increase in gene expression (Figure 4.4E), suggesting that tissue-specific RT and tissue-specific gene expression are mechanistically separable. Therefore, we conclude that differential gene expression between wing discs and follicle cells does not fully explain differences in RT between these two tissues.

As an independent method to assess the relationship between tissue-specific gene expression and RT, we identified genes expressed in both tissues (shared), genes expressed in wing discs only (wing-specific), and genes expressed in follicle cells only (follicle-specific) (Materials and Methods). We identified 12,626 genes that were expressed in both tissues, 901 genes that were wing-specific, and 517 that were follicle-specific (Figure 4.5D). When we quantified differential RT at both shared genes and tissue-specific genes, we observe earlier replication of wing-specific and shared genes in wing discs whereas follicle-specific genes do not replicate earlier in follicle cells. These data again indicate that tissue-specific transcription and tissue-specific RT, although correlated, are separable (Figure 4.5E,F). We hypothesized that earlier replication of shared genes in wing discs would correlate with elevated gene expression genome-wide in wing discs relative to follicle cells. Direct comparison of gene expression between the two tissues revealed a global increase of transcript abundance in wing discs relative to follicle cells (Figure 4.4F,G). Together, these data demonstrate that while gene expression and RT are correlated genome-wide (Figure

4.4H,I), changes in gene expression do not direct changes in RT between wing discs and follicle cells suggesting that RT and transcriptional activity are mechanistically separable.

Figure 4.3. Replicate correlations of RNA-seq data. A) Heatscatter plot comparing wild type follicle cell RNA-seq transcript per million (TPM) values (log_2; top) and *Rif1*$^-$ follicle cell RNA-seq transcript per million (TPM) values (log_2; bottom). **B)** Heatscatter plot comparing wild type wing disc RNA-seq transcript per million (TPM) values (log_2; top) and *Rif1*$^-$ wing disc RNA-seq transcript per million (TPM) values (log_2; bottom). Experiments were performed in collaboration with Souradip Das.

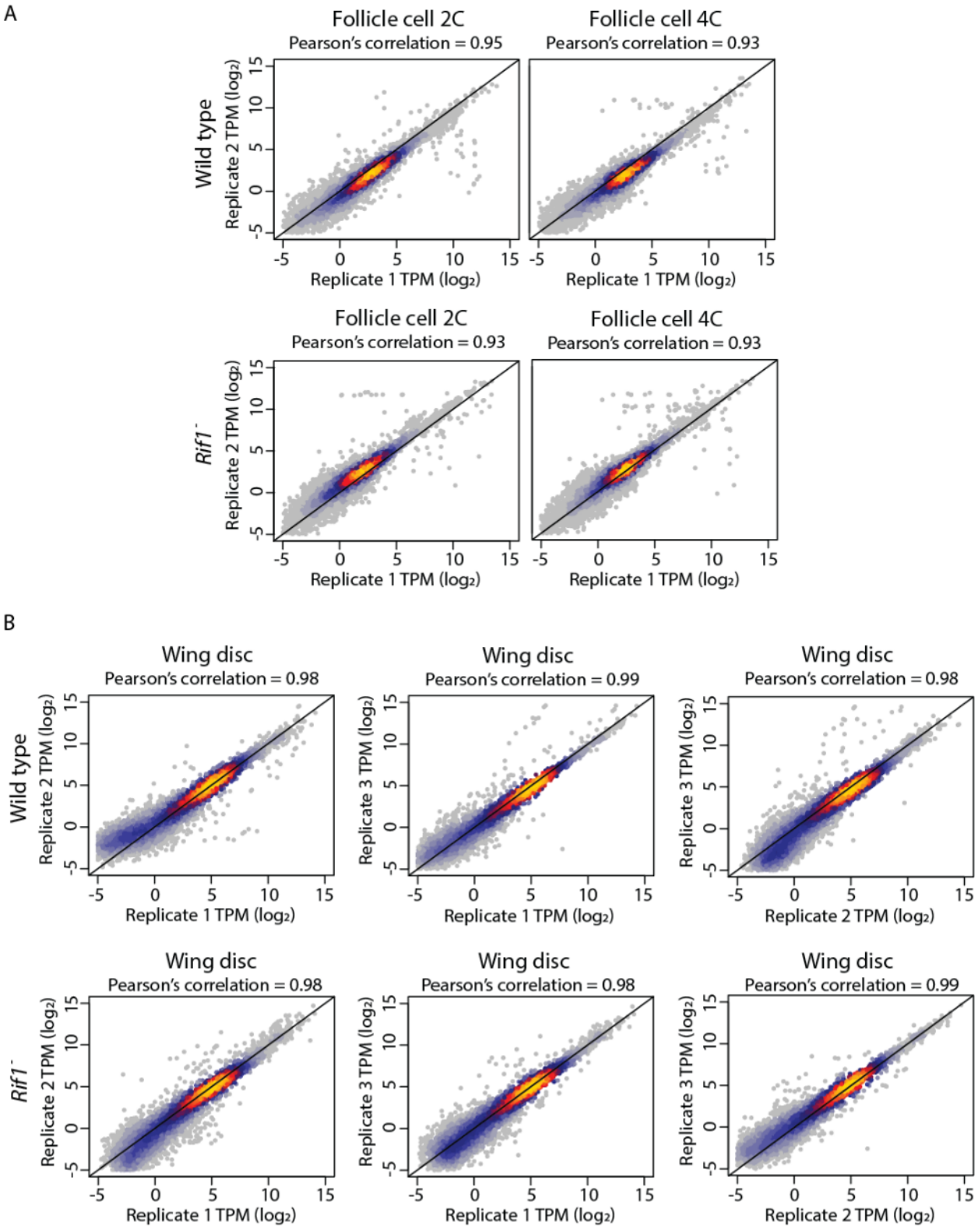

A
Follicle cell 2C
Pearson's correlation = 0.95
Follicle cell 4C
Pearson's correlation = 0.93
Wild type
Follicle cell 2C
Pearson's correlation = 0.93
Follicle cell 4C
Pearson's correlation = 0.93
Rif1-
Replicate 2 TPM (log2)
Replicate 1 TPM (log2)
B
Wing disc
Pearson's correlation = 0.98
Wing disc
Pearson's correlation = 0.99
Wing disc
Pearson's correlation = 0.98
Wild type
Wing disc
Pearson's correlation = 0.98
Wing disc
Pearson's correlation = 0.98
Wing disc
Pearson's correlation = 0.99
Rif1-
Replicate 3 TPM (log2)
Replicate 2 TPM (log2)

Figure 4.4. Transcriptional change does not drive differential RT between lineages. A) Genome browser shot of representative lineage-specific genes, wingless (wg) and chorion proteins 18, 15, and 19 (Cp18, Cp15, and Cp19). RNA-seq signal is shown for two replicates of wild type wing discs (green) and wild type follicle cells (blue). **B)** Histogram of the number of transcripts overlapping each 10kb window. Only windows containing at least one transcript are shown. **C)** Heatscatter plot of the wild type follicle cell/wild type wing disc RT values (S/G1 (log_2)) versus the wild type follicle cell/wild type wing disc ratio of the transcriptional change of the most confident transcript (lowest p value) at each window across the major chromosome scaffolds. Only windows containing at least one transcript are shown. **D)** Heatscatter plot of the wild type follicle cell/wild type wing disc RT values (S/G1 (log_2)) versus the wild type follicle cell/wild type wing disc ratio of the transcriptional change of the transcript with the greatest differential expression (absolute maximum log_2 fold-change) at each window across the major chromosome scaffolds. Only windows containing at least one transcript are shown. **E)** Venn diagrams comparing 10kb windows of significantly increased gene expression in wing discs ($p < 0.01$, log_2 fold change < 0; edgeR) to significantly earlier replication in wing discs ($p < 0.05$, log_2 fold change < -0.1; limma) (top; green) and comparing windows of significantly increased gene expression in follicle cells ($p < 0.01$, log_2 fold change > 0; edgeR) to significantly earlier replication in follicle cells ($p < 0.05$, log_2 fold change > 0.1; limma) (bottom; blue). **F)** Heatscatter plot comparing wild type wing disc RNA-seq signal and wild type follicle cell RNA-seq signal. **G)** Quantification of RNA-seq signal. **H)** Heatscatter plot of wild type wing disc (top) and wild type follicle cell (bottom) S/G1 (log_2) replication timing values versus the number of transcripts within 10kb windows across the major chromosome scaffolds. **I)** Heatscatter plot

of wild type wing disc (top) and wild type follicle cell (bottom) S/G1 (log_2) replication timing values versus the average transcriptional activity within 10kb windows across the major chromosome scaffolds. Experiments were performed in collaboration with Souradip Das.

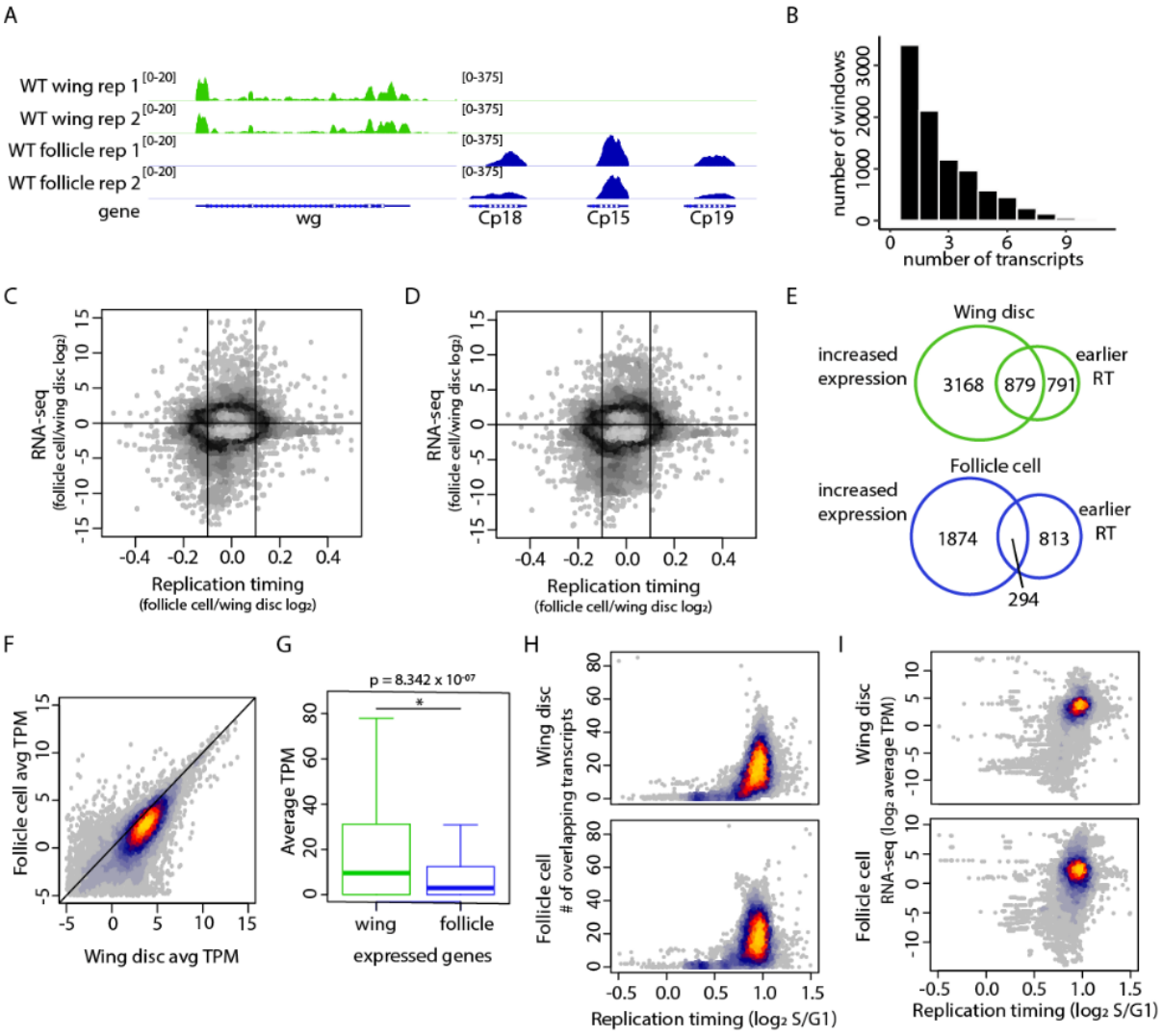
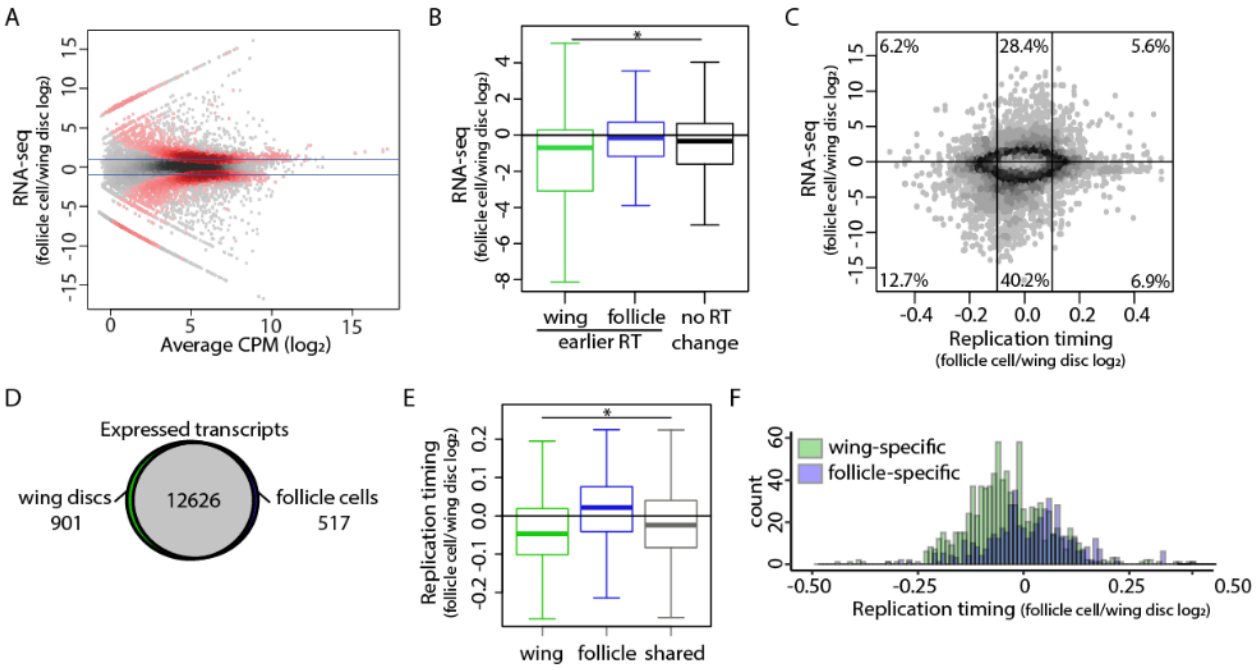


Figure 4.5. Cell type-specific transcription does not drive changes in RT. A) Heatscatter plot of the wild type follicle cell/wild type wing disc ratio of total RNA-seq signal. Statistically different transcripts between wild type follicle cells and wild type wing discs are indicated in red ($p < 0.01$; edgeR). Blue lines indicate a $\log_2$ fold change of 1 and -1. **B)** The average $\log_2$ fold change of all transcripts within each 10kb window of earlier RT in wild type wing discs (green), earlier RT in wild type follicle cells (blue), and unchanged RT (grey). Only windows containing at least one transcript are shown. ($p < 0.0001$; One way ANOVA). **C)** Heatscatter plot of the wild type follicle cell/wild type wing disc RT values (S/G1 ($\log_2$)) versus the wild type follicle cell/wild type wing disc ratio of normalized RNA-seq signal at all 10kb windows across the major chromosome scaffolds. The average $\log_2$ fold change of all transcripts within each 10kb window is plotted, and only windows containing at least one transcript are shown. Percentages represent the number of windows within each region (vertical lines at -0.1 and 0.1 represent $\log_2$ fold change cutoffs for RT statistical significance). **D)** Venn diagram comparing expressed transcripts (TPM > 0) between wild type wing discs and wild type follicle cells. Wing-specific (green), follicle-specific (blue) and shared (grey) transcripts are indicated. **E)** S/G1 ($\log_2$) $\log_2$ fold change between wild type follicle cells and wild type wing discs at wing-specific (green), follicle-specific (blue), and shared (black) transcripts ($p < 0.0001$; One way ANOVA). **F)** Histogram of replication timing $\log_2$ fold change of wing-specific (green) and follicle-specific (blue) transcripts. Experiments were performed in collaboration with Souradip Das.


A
RNA-seq
(follicle cell/wing disc log2)
Average CPM (log2)
B
RNA-seq
(follicle cell/wing disc log2)
wing
follicle
earlier RT
no RT
change
C
6.2%
28.4%
5.6%
12.7%
40.2%
6.9%
RNA-seq
(follicle cell/wing disc log2)
Replication timing
(follicle cell/wing disc log2)
D
Expressed transcripts
wing discs
901
12626
follicle cells
517
E
Replication timing
(follicle cell/wing disc log2)
wing
follicle
shared
F
wing-specific
follicle-specific
count
Replication timing (follicle cell/wing disc log2)

The switch to endoreplication does not affect DNA replication timing in follicle cells

The follicle cells of the adult ovary undergo a developmentally programmed cell cycle transition in which, after a series of mitotic divisions, they begin endocycling, a cell cycle consisting of S and G phases with no intervening mitoses (Figure 4.6A) (Fox and Duronio 2013; Edgar et al. 2014). Follicle cells undergo three endocycles, resulting in a ploidy of 16C. Previous work has shown that there are distinct changes in genome regulation during the endocycle, including a global decrease in transcription, decrease in E2F1 target gene expression, and acquisition of endocycle-specific ORC binding sites (Maqbool et al. 2010; Sher et al. 2012; Hua et al. 2018; Rotelli et al. 2019). Therefore, we hypothesized that follicle cell replication timing may be influenced by this developmentally regulated cell cycle transition.

To determine if the transition from a mitotic cycle to an endocycle causes a change in RT, we generated genome-wide replication timing profiles from wild type endocycling follicle cells and compared them to the RT profiles we measured from wild type mitotic follicle cells (Figure 4.7A,B). To this end, we collected the S phase populations between the 2C and 4C peaks (mitotic) and between the 4C and 8C peaks, which corresponds to the second of the three endocycles (Figure 4.6B). Direct comparison of RT profiles generated from wild type mitotic (2C-4C) and endocycling (4C-8C) follicle cells showed no windows of differential RT genome-wide between the two populations of follicle cells (Figure 4.6C; Figure 4.7C; Table 4.1). Likewise, the gene expression profiles of these two populations of follicle cells were highly similar, with only six differentially expressed transcripts between mitotically cycling and endocycling follicle cells ($p < 0.01$, edgeR; Figure 4.3; Figure 4.7D). It is important to note that the first follicle cell endocycle likely initiates from G1 phase (Lilly

and Spradling 1996; Calvi et al. 1998); therefore, the mitotic S phase sample may contain both mitotic and endocycling follicle cells. We were concerned that the impure cell population in the mitotic follicle cell dataset might mask any differential RT between the mitotic and endocycling populations. Based on the number of follicle cells in a mature egg chamber (~1000), we estimate that follicle cells in the first endo S phase could account for, at most, one half of the 'mitotic' follicle cell population (2C-4C) (Materials and Methods). Therefore, we performed an *in silico* false discovery rate (FDR) analysis by spiking in random reads from the wing disc RT dataset into the mitotic follicle cell RT dataset. Given that the endocycling follicle cells contribute no more than 50% of our total mitotic follicle cell population, our analysis would be sensitive enough to accurately identify at least ~27% of the endocycle-specific RT differences (Figure 4.7E; Materials and Methods). Thus, endo S cells in the 2C-4C population do not mask a difference in RT between endocycling and mitotic follicle cells. Although we cannot exclude the possibility that minor changes in RT could be masked in in our data. we conclude that mitotic and endocycling follicle cells have remarkably similar RT profiles, arguing that cell lineage, not changes in the cell cycle is a major contributing factor to RT.

Figure 4.6. S phase strategy does not affect DNA replication timing within the follicle cells of the adult ovary.

A) Early egg chamber development within the adult *Drosophila* ovary. **B)** Representative FACS profile of follicle cell nuclei isolated from whole ovaries. The 2C-4C S phase fraction (blue) are the mitotically cycling follicle cells, and the 4C-8C S phase fraction (orange) are the endocycling follicle cells. **C)** LOESS regression line showing average wild type mitotically cycling follicle cells (blue) and wild type endocycling follicle cells (orange) S/G1 (log_2) replication timing values in at across the chromosome 3L scaffold. See Figure 4.7 for all other chromosome arms. **D)** Correlation matrix of S/G1 (log_2) replication timing values for wild type endocycling follicle cells (endo S), wild type mitotically cycling follicle cells (mitotic S), and wild type wing discs. Experiments were performed in collaboration with Souradip Das.

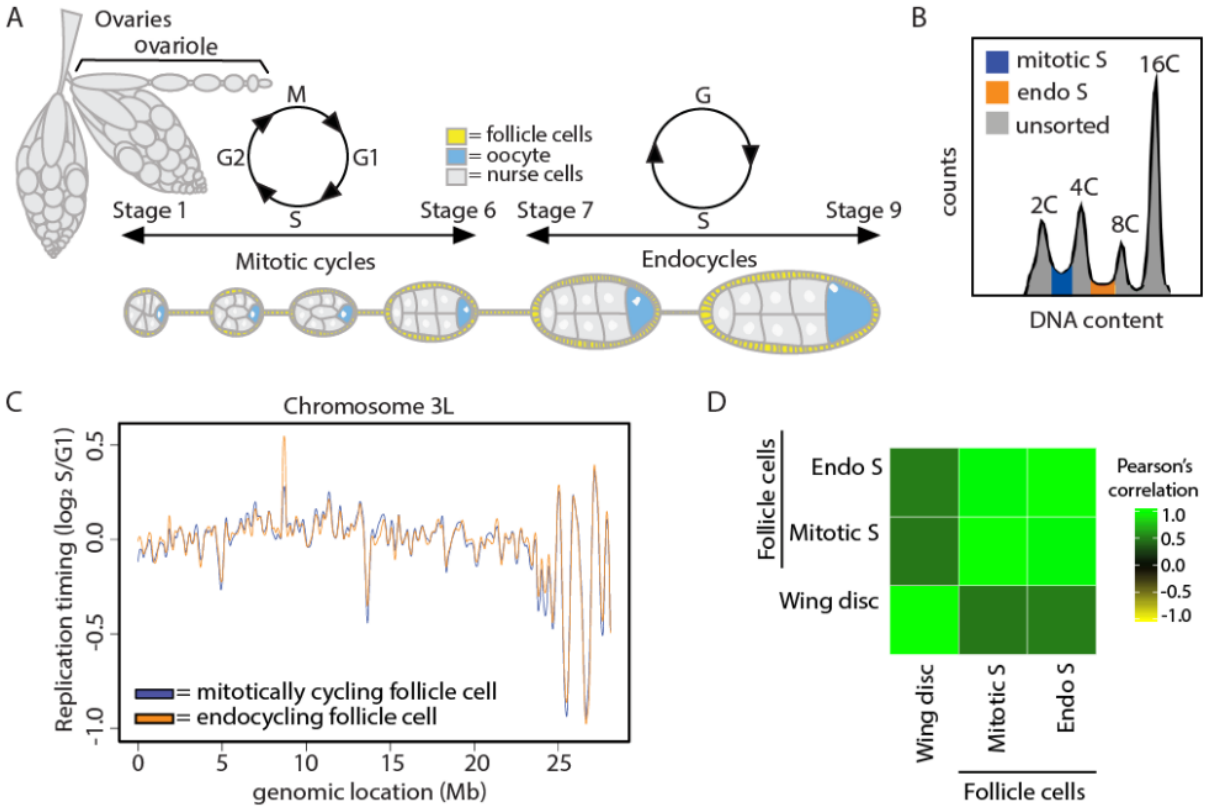
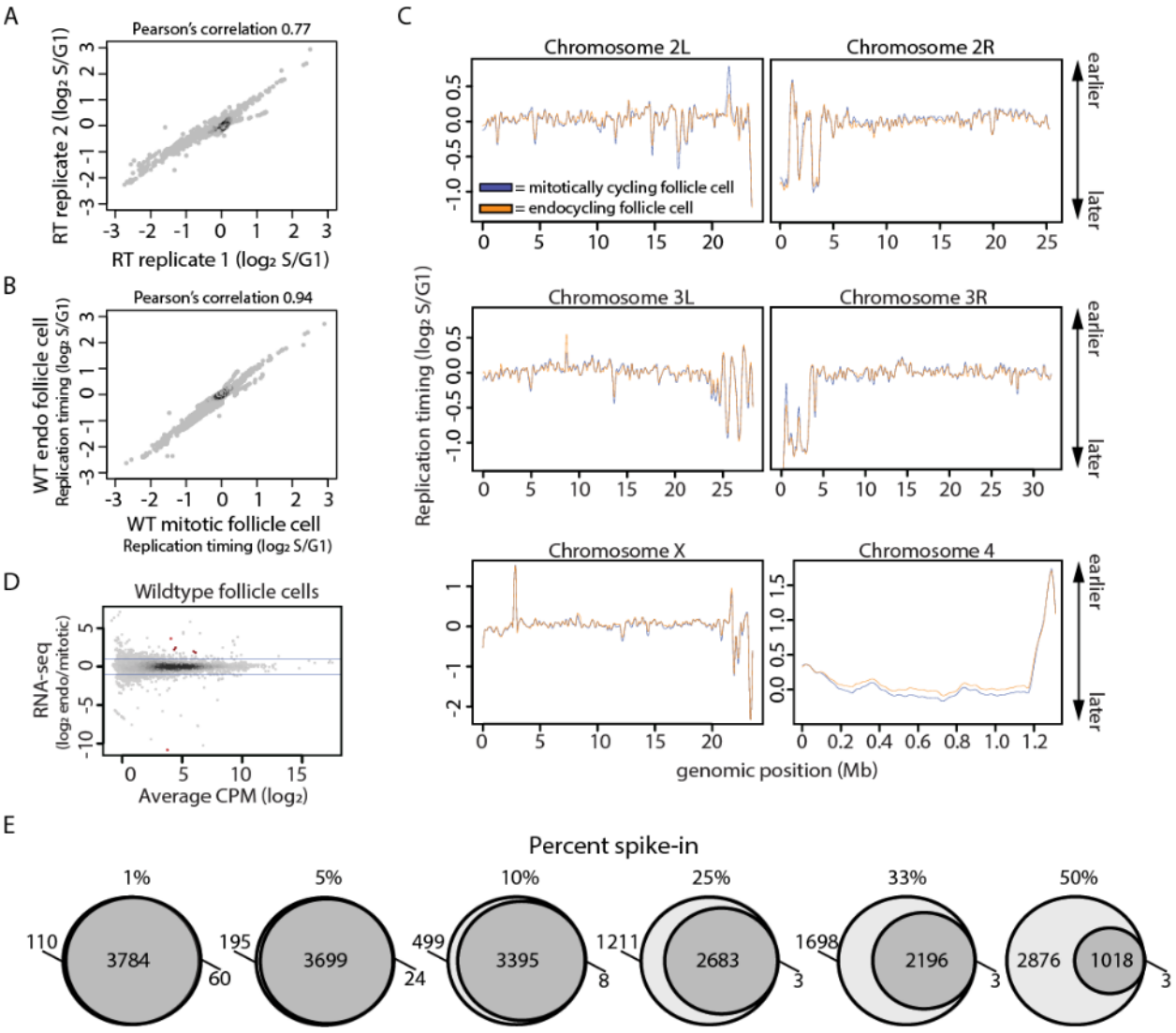


Figure 4.7. Characterization of RT between wild type mitotically cycling and endocycling follicle cells.

A) Heatscatter plot comparing wild type endocycling follicle cell S/G1 (log_2) replicate replication timing values. **B)** Heatscatter plot comparing wild type mitotically cycling follicle cell and endocycling follicle cells S/G1 (log_2) ratios at all 100kb windows using a 10kb slide across all major chromosome scaffolds. **C)** LOESS regression line showing average S/G1 (log_2) replication timing values for wild type mitotically cycling follicle cells (blue) and wild type endocycling follicle cells (orange) at 100kb windows using a 10kb slide across the major chromosome scaffolds. **D)** Heatscatter plot of the wild type endocycling follicle cell/wild type mitotically cycling follicle cell ratio of total RNA-seq signal. Statistically different transcripts between wild type follicle cells and wild type wing discs are indicated in red ($p < 0.01$; edgeR). Blue lines indicate a log_2 fold change of 1 and -1. **E)** Venn diagram comparisons of significant RT changes identified between *in* silico-generated spike-in datasets (Materials and methods) and wild type wing imaginal discs (dark grey) versus significant RT changes identified between wild type follicle cells and wild type wing imaginal discs (light grey; $p < 0.01$, absolute log_2 fold change > 0.1; limma). Experiments were performed in collaboration with Souradip Das.


A
Pearson's correlation 0.77
RT replicate 2 (log2 S/G1)
RT replicate 1 (log2 S/G1)
B
Pearson's correlation 0.94
WT endo follicle cell
Replication timing (log2 S/G1)
WT mitotic follicle cell
Replication timing (log2 S/G1)
C
Chromosome 2L
Chromosome 2R
= mitotically cycling follicle cell
= endocycling follicle cell
Chromosome 3L
Chromosome 3R
Chromosome X
Chromosome 4
earlier
later
Replication timing (log2 S/G1)
genomic position (Mb)
D
Wildtype follicle cells
RNA-seq
(log2 endo/mitotic)
Average CPM (log2)
E
Percent spike-in
1%
5%
10%
25%
33%
50%
110
3784
60
195
3699
24
499
3395
8
1211
2683
3
1698
2196
3
2876
1018
3

Rif1 fine tunes the replication timing program in different tissues

Rif1 is a global regulator of DNA RT from yeast to humans (Cornacchia et al. 2012; Hayano et al. 2012; Yamazaki et al. 2012; Peace et al. 2014; Seller and O'Farrell 2018). We sought to determine whether Rif1 regulates RT in a tissue-specific manner or whether Rif1-dependent RT domains are hardwired into the genome. To address these questions, we generated genome-wide RT profiles from mitotic follicle cells and wing discs in a *Rif1* null (*Rif1*$^{-}$) mutant previously generated by our lab (Figure 4.8A,B; (Munden et al. 2018)). Individual replicates of *Rif1*$^{-}$ RT data generated from either wing discs or follicle cells correlated well (Figure 4.8C; Figure 4.9A), whereas comparison of *Rif1*$^{-}$ and wild type RT data revealed that approximately 13% of the genome has differential RT in mitotically cycling follicle cells and 8% of the genome has differential RT in wing discs (Pearson's correlation coefficient = 0.52 and 0.78, respectively; Figure 4.9B; Figure 4.8D). For the *Rif1*$^{-}$ mutant follicle cells, 8.2% of windows displayed advanced RT while 5.0% of windows had delayed RT (Figure 4.10A-C; Figure 4.9C; Table 4.1). In the *Rif1*$^{-}$ mutant wing disc, 4.1% of windows had advanced RT and 3.9% of windows had delayed RT (Figure 4.10A-C; Figure 4.8E; Table 4.1). Furthermore, the magnitude of RT changes within windows of differential RT between *Rif1*$^{-}$ and wild type was significantly greater in follicle cells than that observed in wing discs (Figure 4.10B,D). These data show that Rif1 has a greater impact on RT in follicle cells than wing discs, arguing that Rif1-dependent RT domains are not hardwired into the genome.

Figure 4.8. Characterization of RT in *Rif1*$^{-}$ wing imaginal discs. A) Western blot analysis of protein isolated from 10, 20, and 40 wild type and *Rif1*$^{-}$ wing discs (left to right). **B)** Wild type and *Rif1*$^{-}$ wing imaginal disc cells stained with DAPI (blue) and anti-Rif1 (green) antibodies. Bar, 10 μm. **C)** Heatscatter plot comparing *Rif1*$^{-}$ wing disc S/G1 (log_2) replicate replication timing values. **D)** Heatscatter plot comparing wild type and *Rif1*$^{-}$ wing disc S/G1 (log_2) ratios at 100kb windows using a 10kb slide across all major chromosome scaffolds. **E)** LOESS regression line showing average S/G1 (log_2) replication timing values for wild type wing discs (black) and *Rif1*$^{-}$ wing discs (cyan) at 100kb windows using a 10kb slide across the major chromosome scaffolds. Experiments were performed in collaboration with Christina Hill.

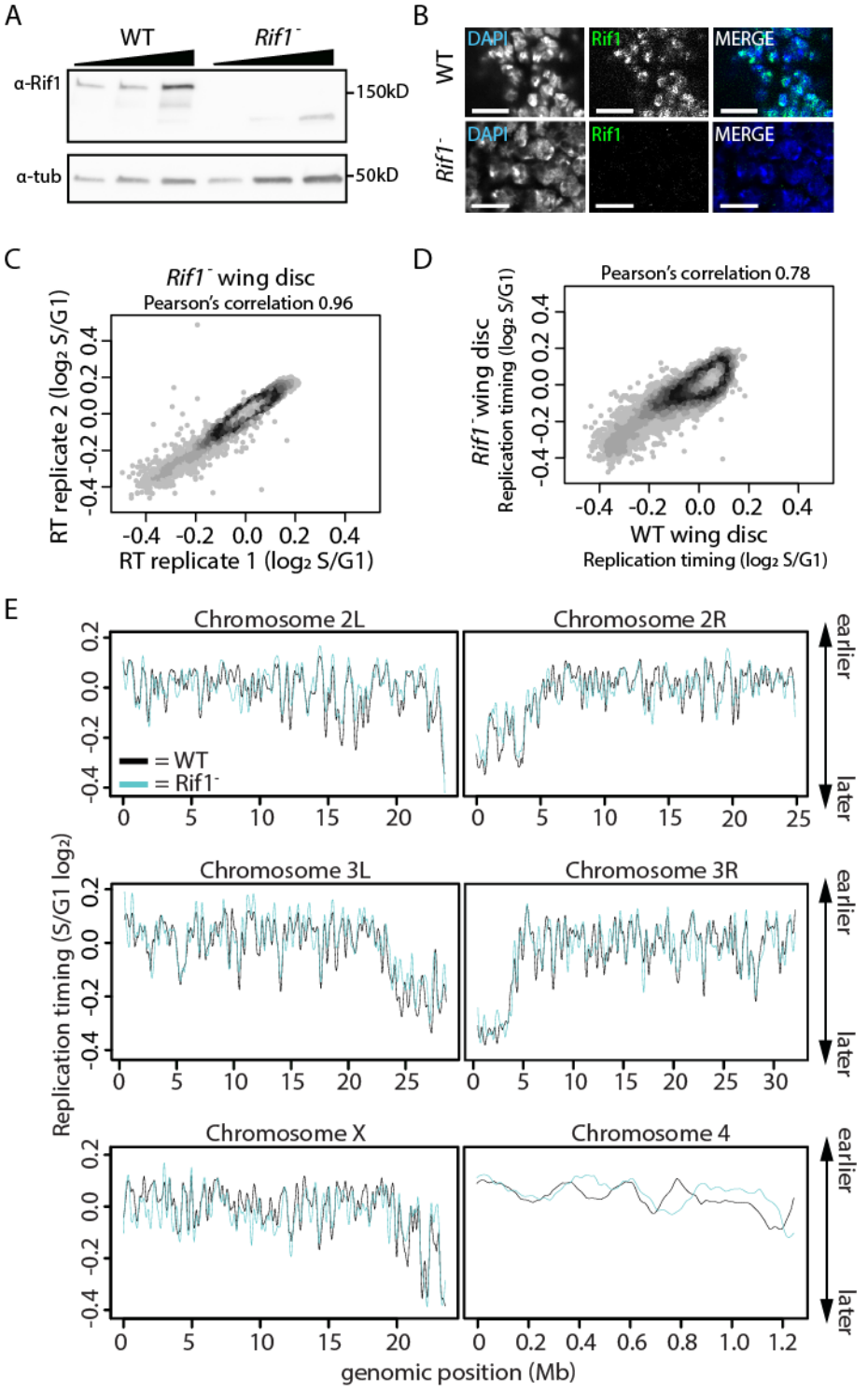

A
WT
Rif1-
α-Rif1
150kD
α-tub
50kD
B
WT
Rif1-
DAPI
Rif1
MERGE
C
Rif1- wing disc
Pearson's correlation 0.96
RT replicate 2 (log2 S/G1)
RT replicate 1 (log2 S/G1)
D
Pearson's correlation 0.78
Rif1- wing disc
Replication timing (log2 S/G1)
WT wing disc
Replication timing (log2 S/G1)
E
Chromosome 2L
Chromosome 2R
Chromosome 3L
Chromosome 3R
Chromosome X
Chromosome 4
= WT
= Rif1-
earlier
later
Replication timing (S/G1 log2)
genomic position (Mb)

Figure 4.9. Characterization of RT in *Rif1*$^{-}$ mitotically cycling follicle cells. A) Heatscatter plot comparing *Rif1*$^{-}$ mitotically cycling follicle cell S/G1 (log_2) replicate replication timing values. **B)** Heatscatter plot comparing wild type and *Rif1*$^{-}$ mitotically cycling follicle cell S/G1 (log_2) ratios at 100kb windows using a 10kb slide across all major chromosome scaffolds. C**)** LOESS regression line showing average S/G1 (log_2) replication timing values for wild type mitotically cycling follicle cells (black) and *Rif1*$^{-}$ mitotically cycling follicle cells (cyan) at 100kb windows using a 10kb slide across the major chromosome scaffolds. **D)** Heatscatter plot of the *Rif1*$^{-}$/control ratio of total RNA-seq signal in follicle cells (top) and wing discs (bottom). Statistically different transcripts are indicated in red ($p < 0.01$; edgeR). Blue lines indicate a log_2 fold change of 1 and -1. Experiments were performed in collaboration with Souradip Das.

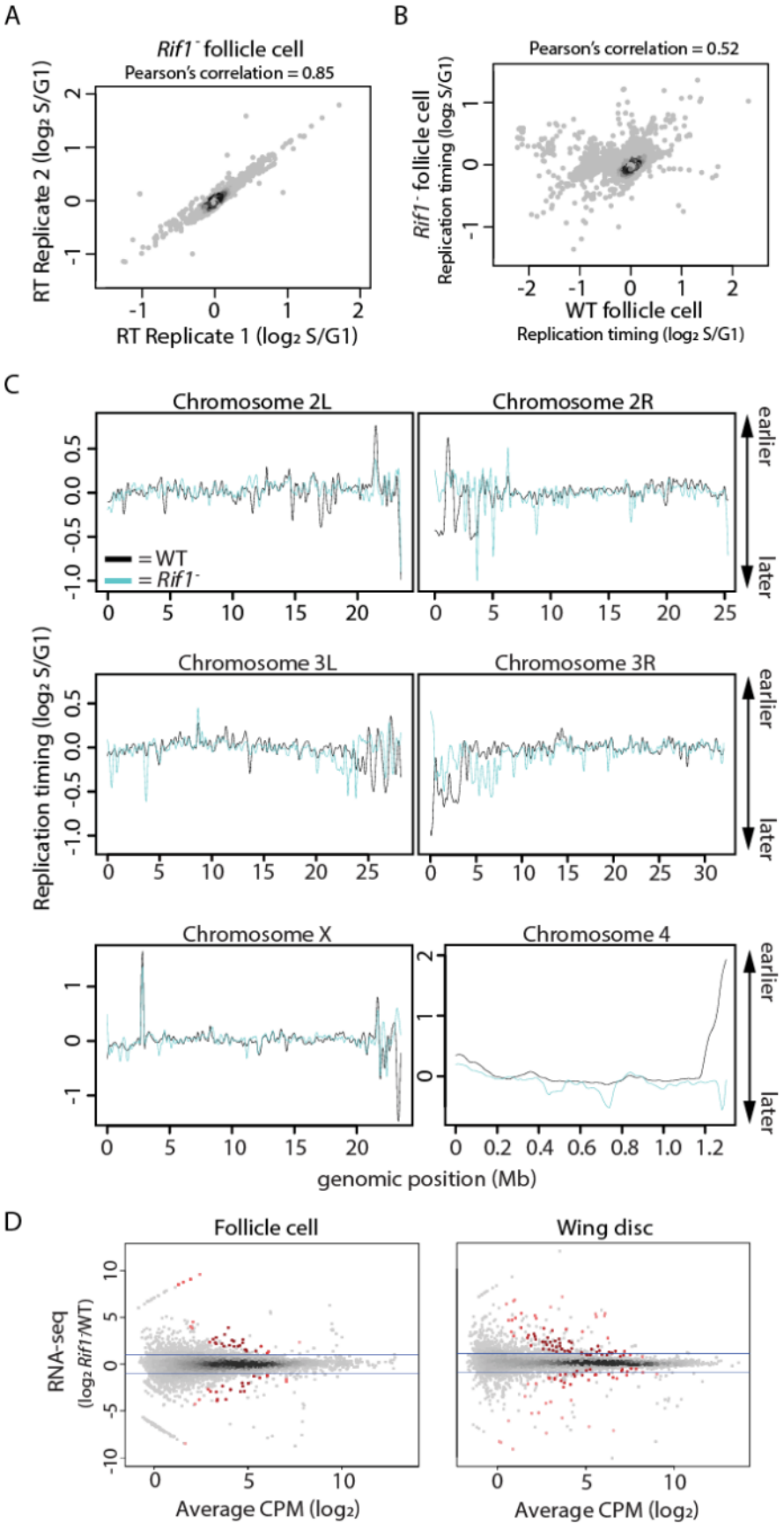
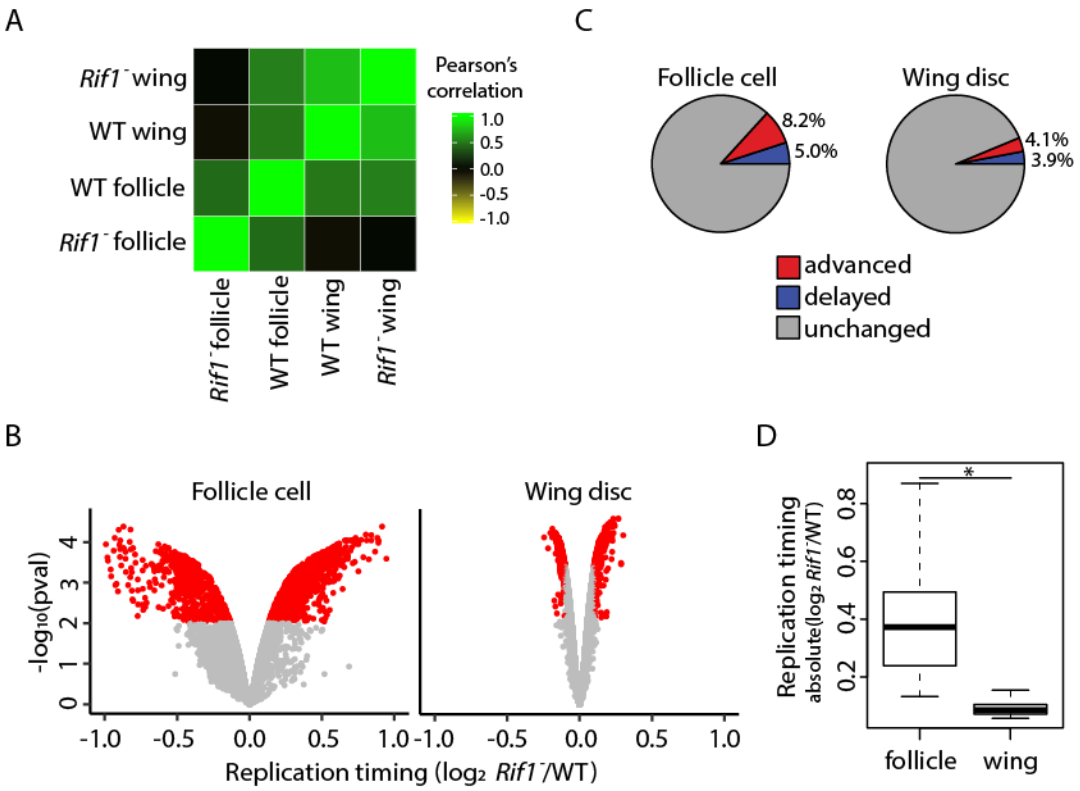

A
Rif1- follicle cell
Pearson's correlation = 0.85
RT Replicate 2 (log2 S/G1)
RT Replicate 1 (log2 S/G1)
B
Pearson's correlation = 0.52
Rif1- follicle cell
Replication timing (log2 S/G1)
WT follicle cell
Replication timing (log2 S/G1)
C
Chromosome 2L
Chromosome 2R
= WT
= Rif1-
earlier
later
Chromosome 3L
Chromosome 3R
Replication timing (log2 S/G1)
Chromosome X
Chromosome 4
genomic position (Mb)
D
Follicle cell
Wing disc
RNA-seq
(log2 Rif1/WT)
Average CPM (log2)
Average CPM (log2)

Figure 4.10. Rif1 regulates RT in a lineage-specific manner. A) Correlation matrix of S/G1 (log_2) replication timing values for wild type mitotically cycling follicle cells (WT follicle), *Rif1*$^-$ mitotically cycling follicle cells (*Rif1*$^-$ follicle), wild type wing discs (WT wing), and *Rif1*$^-$ wing discs (*Rif1*$^-$ wing). **B)** Volcano plot of the *Rif1*$^-$/control ratio of normalized replication timing values (S/G1 (log_2)) plotted versus the -log_{10} p value (adjusted for multiple testing) in follicle cells (left) and wing discs (right). Significant replication timing changes are indicated (red; $p < 0.01$, absolute log_2 fold change > 0.1; limma). **C)** Pie chart of all 100kb windows of significantly advanced RT (red), significantly delayed RT (blue), and unchanged RT (grey) across the major chromosome scaffolds in *Rif1*$^-$ mutants relative to wild type control in follicle cells (left) and wing discs (right) **D)** S/G1 (log_2) absolute log_2 fold change at 100kb windows of significant RT change between *Rif1*$^-$ and control in follicle cells and wing discs (Student's t test, $p < 2.2 \times 10^{-16}$). Experiments were performed in collaboration with Souradip Das.


A
Rif1- wing
WT wing
WT follicle
Rif1- follicle
Rif1- follicle
WT follicle
WT wing
Rif1- wing
Pearson's correlation
1.0
0.5
0.0
-0.5
-1.0
C
Follicle cell
Wing disc
8.2%
5.0%
4.1%
3.9%
advanced
delayed
unchanged
B
Follicle cell
Wing disc
-log10(pval)
0
1
2
3
4
-1.0
-0.5
0.0
0.5
1.0
Replication timing (log2 Rif1-/WT)
D
Replication timing
absolute(log2 Rif1/WT)
0.2
0.4
0.6
0.8
*
follicle
wing

Rif1 promotes late replication likely by preventing replicative helicase activation (Hayano et al. 2012; Davé et al. 2014; Hiraga et al. 2014; Mattarocci et al. 2014; Hiraga et al. 2017). Therefore, we hypothesized that advanced RT in a *Rif1*$^{-}$ mutant is a direct effect of loss of Rif1 function, whereas delayed RT in a *Rif1*$^{-}$ mutant is a secondary effect. This hypothesis predicts that when comparing different *Rif1*$^{-}$ mutant cell types there should be a greater extent of overlap between regions with advanced RT (direct) than between regions with delayed RT (indirect) in *Rif1*$^{-}$ mutants. We found that 43.8% (242/552) of windows with advanced RT in wing discs were also advanced in follicle cells. In contrast, only 16.9% (89/527) of windows with delayed RT in wing discs were also delayed in follicle cells (Figure 4.11A). These data support the hypothesis that advanced RT is a direct effect of Rif1 loss whereas delayed RT may be a secondary effect.

While measuring RT values for *Rif1* mutant and control samples, we profiled *Rif1*$^{-/+}$ heterozygous follicle cells (Figure 4.12A,B). To our surprise, this heterozygous genotype displayed an intermediate RT phenotype with 3.6% (478/13391) of windows with advanced RT and 1.6% of windows with delayed RT relative to wild type follicle cells (Figure 4.12C). Furthermore, 87.0% of windows with significantly advance and 57.5% with significantly delayed RT in *Rif1* heterozygotes were also affected in *Rif1*$^{-}$ follicle cells, indicating dependency on Rif1 function (Figure 4.12D). These data demonstrate that Rif1 is haploinsufficient for RT control in follicle cells.

As an independent metric to address the specificity of commonly advanced and/or delayed RT changes, we asked whether common RT changes between mitotic follicle cells and wing discs were also detected in *Rif1*$^{-}$ endocycling follicle cells. We generated RT profiles from *Rif1*$^{-}$ endocycling follicle cells and found that individual replicates of RT data

correlated well (Figure 4.13A). In contrast, 14.8% of windows displayed differential RT in *Rif1*$^{-}$ endocycling follicle cells relative to control with 7.2% being advanced and 7.6% being delayed (Figure 4.11B; Figure 4.13B; Table 4.1). Although RT was similar between wild type mitotic and endocycling follicles cells, *Rif1* mutation affected these cell populations differently. We found that 72.1% (789/960) of advanced windows in *Rif1*$^{-}$ endocycling follicle cells were also advanced in *Rif1*$^{-}$ mitotic follicle cells, and only 37.9% (388/1024) of the windows that were delayed in *Rif1*$^{-}$ endocycling follicle cells were also delayed in *Rif1*$^{-}$ mitotic follicle cells (Figure S8C). Accordingly, the low degree of overlap between windows of delayed RT is reflected by the low genome-wide RT correlation between *Rif1*$^{-}$ mitotic and endocycling follicle cells (Figure 4.11B; Figure 4.13D). Interestingly, many of the regions of advanced RT changes that were in common between *Rif1*$^{-}$ wing discs and mitotic follicle cells were also detected in *Rif1*$^{-}$ endocycling follicle cells while the delayed RT changes were mostly non-overlapping (72.7% (176/242) and 47.2% (42/89), respectively). Therefore, while Rif1 regulates RT in a tissue-specific manner, Rif1 appears to regulate RT in a core region of the genome regardless of cell type.

Figure 4.11. Rif1 promotes late replication of pericentric heterochromatin across lineages.

A) Venn diagrams comparing significantly advanced (top) and delayed (bottom) 100kb windows identified in *Rif1*$^{-}$ follicle cells (left; blue) and wing discs (right; green) ($p<0.01$ and absolute $\log_2$ fold change > 0.1; limma). **B)** Correlation matrix of S/G1 ($\log_2$) replication timing values for wild type mitotically cycling follicle cells (WT mitotic S), *Rif1*$^{-}$ mitotically cycling follicle cells (*Rif1*$^{-}$ mitotic S), wild type endocycling follicle cells (WT endo S), *Rif1*$^{-}$ mitotically cycling follicle cells (*Rif1*$^{-}$ endo S), wild type wing discs (WT wing), and *Rif1*$^{-}$ wing discs (*Rif1*$^{-}$ wing). **C)** Pie chart of all 100kb windows of commonly advanced RT between *Rif1*- wing discs and follicle cells. Windows within pericentromeres are in grey and chromosome arms are in black. **D)** Bar plot of the percentage of 100kb windows in pericentric heterochromatin with significantly advanced RT. **E)** S/G1 ($\log_2$) absolute $\log_2$ fold change at all 100kb windows located in pericentric heterochromatin between *Rif1*$^{-}$ and control (Student's t test, $p < 2.2 \times 10^{-16}$). **F)** Heatscatter plot of the *Rif1*$^{-}$/control ratio of normalized replication timing values (S/G1 ($\log_2$)) plotted versus the *Rif1*$^{-}$/control ratio of the most confident transcript (lowest p value) at each window across the major chromosome scaffolds. Significantly advanced (red) and delayed (blue) windows are indicated ($p < 0.05$, absolute $\log_2$ fold change > 0.1 (vertical lines); limma). Experiments were performed in collaboration with Souradip Das.

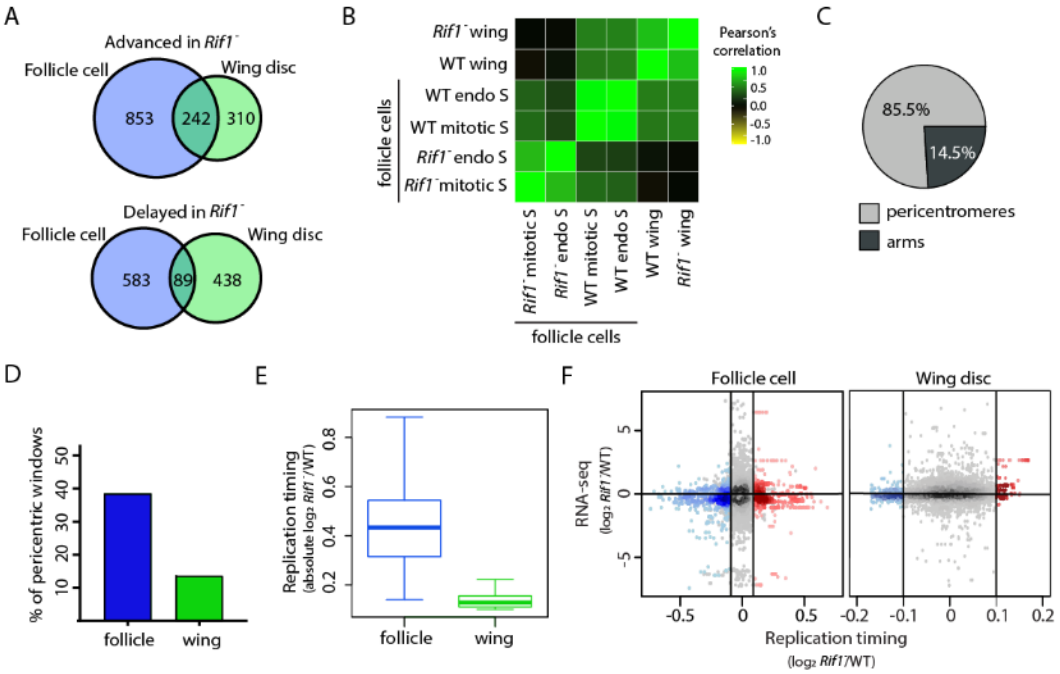

A
Advanced in Rif1-
Follicle cell
Wing disc
853
242
310
Delayed in Rif1-
Follicle cell
Wing disc
583
89
438
B
Rif1- wing
WT wing
WT endo S
WT mitotic S
Rif1- endo S
Rif1- mitotic S
follicle cells
Pearson's correlation
1.0
0.5
0.0
-0.5
-1.0
Rif1- mitotic S
Rif1- endo S
WT mitotic S
WT endo S
WT wing
Rif1- wing
follicle cells
C
85.5%
14.5%
pericentromeres
arms
D
% of pericentric windows
10
20
30
40
50
follicle
wing
E
Replication timing
(absolute log2 Rif1-/WT)
0.2
0.4
0.6
0.8
follicle
wing
F
Follicle cell
Wing disc
RNA-seq
(log2 Rif1-/WT)
5
0
-5
-0.5
0
0.5
-0.2
-0.1
0
0.1
0.2
Replication timing
(log2 Rif1-/WT)

Figure 4.12. Characterization of RT in *Rif1*$^{-/+}$ mitotic follicle cells. **A)** Heatscatter plot comparing *Rif1*$^{-/+}$ mitotic follicle cell S/G1 (log_2) replicate replication timing values. **B)** LOESS regression line showing average S/G1 (log_2) replication timing values for wild type mitotic follicle cells (black) and *Rif1*$^{-/+}$ mitotic follicle cells (light green) at 100kb windows using a 10kb slide across the major chromosome scaffolds. **C)** Pie chart of all 100kb windows of significantly advanced (red), delayed (blue), and unchanged RT (grey) in *Rif1*$^{-/+}$ mitotic follicle cells across the major chromosome scaffolds. **D)** Venn diagrams comparing significantly advanced (left) and delayed (right) 100kb windows identified in *Rif1*$^{-}$ and *Rif1*$^{-/+}$ follicle cells ($p<0.01$ and absolute log_2 fold change > 0.1; limma). Experiments were performed in collaboration with Souradip Das.

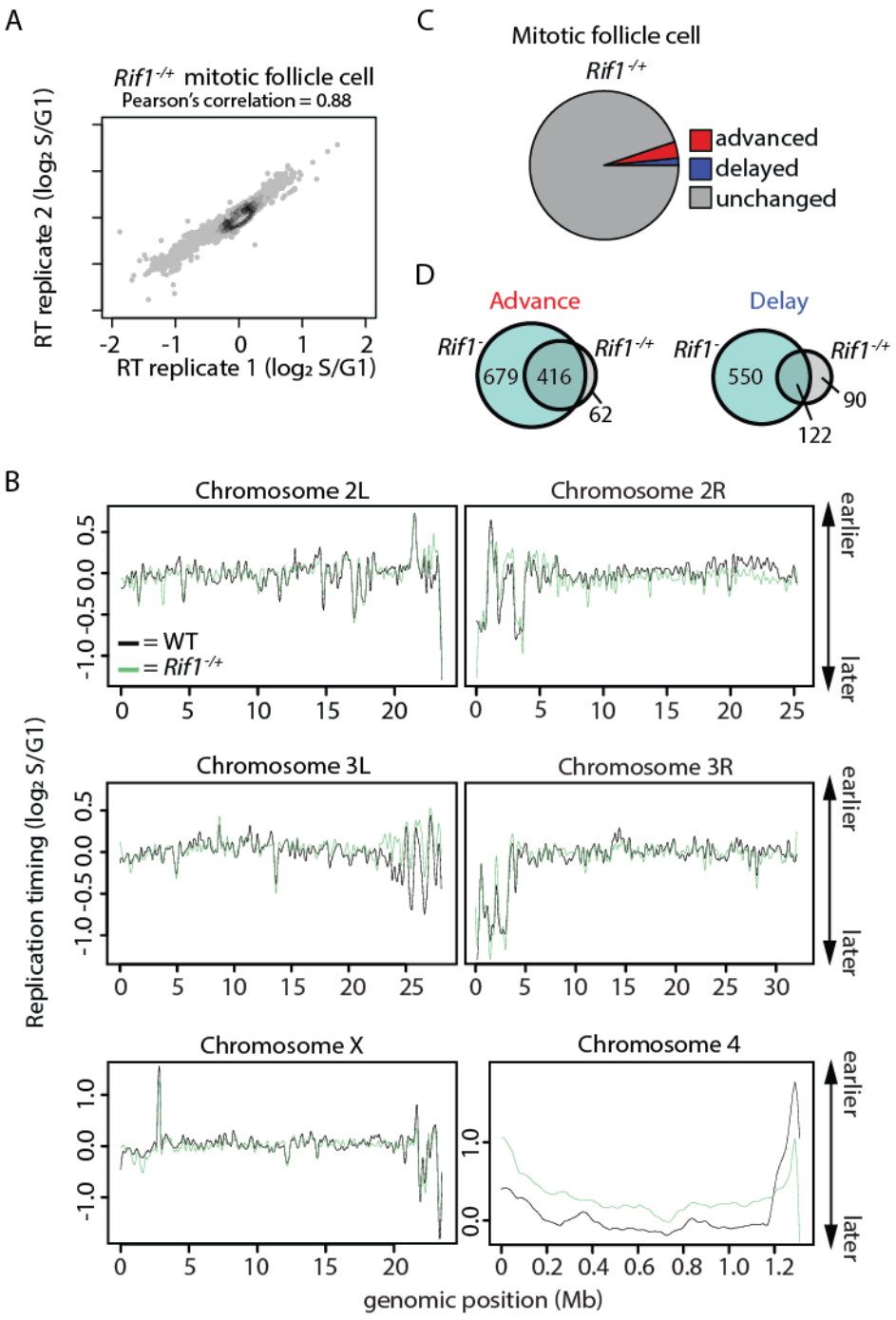
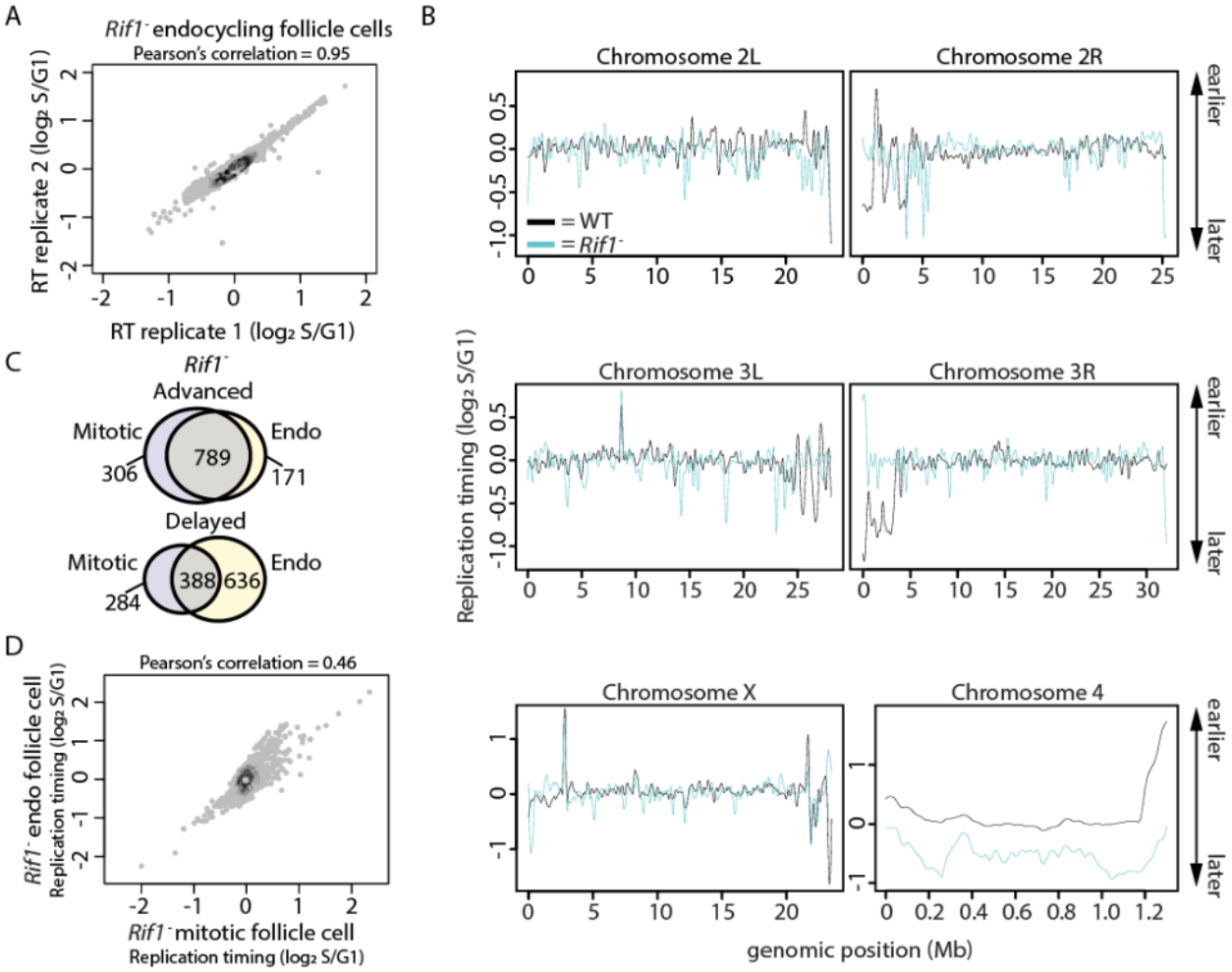

A
Rif1-/+ mitotic follicle cell
Pearson's correlation = 0.88
RT replicate 2 (log2 S/G1)
RT replicate 1 (log2 S/G1)
C
Mitotic follicle cell
Rif1-/+
advanced
delayed
unchanged
D
Advance
Delay
Rif1-
Rif1-/+
679
416
62
550
90
122
B
Chromosome 2L
Chromosome 2R
Chromosome 3L
Chromosome 3R
Chromosome X
Chromosome 4
= WT
= Rif1-/+
earlier
later
Replication timing (log2 S/G1)
genomic position (Mb)

Figure 4.13. Characterization of RT in *Rif1*$^{-}$ endocycling cycling follicle cells. A) Heatscatter plot comparing *Rif1*$^{-}$ endocycling follicle cell S/G1 (log_2) replicate replication timing values. **B)** LOESS regression line showing average S/G1 (log_2) replication timing values for wild type mitotically cycling follicle cells (black) and *Rif1*$^{-}$ mitotically cycling follicle cells (cyan) at 100kb windows using a 10kb slide across the major chromosome scaffolds. **C)** Venn diagrams comparing significant advanced (top) and significantly delayed (bottom) RT changes identified in *Rif1*$^{-}$ mitotically cycling follicle cells (left) and *Rif1*$^{-}$ endocycling follicle cells (right). **D)** Heatscatter plot comparing *Rif1*$^{-}$ mitotically cycling and endocycling follicle cell S/G1 (log_2) ratios at 100kb windows using a 10kb slide across all major chromosome scaffolds. Experiments were performed in collaboration with Souradip Das.


A
Rif1- endocycling follicle cells
Pearson's correlation = 0.95
RT replicate 2 (log2 S/G1)
RT replicate 1 (log2 S/G1)
B
Chromosome 2L
Chromosome 2R
Chromosome 3L
Chromosome 3R
Chromosome X
Chromosome 4
= WT
= Rif1-
earlier
later
Replication timing (log2 S/G1)
genomic position (Mb)
C
Rif1-
Advanced
Mitotic
Endo
306
789
171
Delayed
Mitotic
Endo
284
388
636
D
Pearson's correlation = 0.46
Rif1- endo follicle cell
Replication timing (log2 S/G1)
Rif1- mitotic follicle cell
Replication timing (log2 S/G1)

Rif1 controls RT of pericentric heterochromatin

Almost all commonly advanced windows in *Rif1*$^{-}$ mutant cell populations are located within pericentric heterochromatin, where Rif1 is known to be located (Buonomo et al. 2009; Munden et al. 2018; Seller and O'Farrell 2018). In contrast, all but eight of the commonly delayed windows are located along euchromatic chromosome arms (Figure 4.11C; Figure 4.14A). This relationship is also true for tissue-specific RT changes in *Rif1*$^{-}$ wing discs and follicle cells—advancements are over-represented in pericentric heterochromatin whereas delays are over-represented along chromosome arms (Figure 4.14B). Collectively, these data suggest that Rif1 directly regulates late replication and may play a significant role in regulating late replication of pericentric heterochromatin. Interestingly, almost 40% of pericentric heterochromatin advances in *Rif1*$^{-}$ follicle cells (both mitotically cycling and endocycling), whereas 2.8-fold fewer pericentric windows advance RT in *Rif1*$^{-}$ wing discs (Figure 4.11D; Figure 4.14B). Furthermore, the overall RT of *Rif1*$^{-}$ pericentric heterochromatin remains very late in wing discs relative to the average RT of the chromosome arms, and the magnitude of RT advancement is less than that observed in *Rif1*$^{-}$ pericentric heterochromatin in follicle cells (Figure 4.11E; Figure 4.8E). Therefore, Rif1 contributes more substantially to late replication of pericentric heterochromatin in follicle cells than in wing discs.

Some regions of *Drosophila* polyploid genomes are under-replicated relative to the rest of the genome; i.e. they have endoreplicated less extensively. This is particularly true in pericentric heterochromatin in salivary glands, and this under-replication requires Rif1 (Munden et al. 2018). Consequently, because our RT protocol measures relative copy number in S phase versus G1 phase, one possible explanation for the significantly earlier

replication of pericentric heterochromatin in polyploid *Rif1*$^{-}$ follicle cells relative to diploid *Rif1*$^{-}$ wing discs is a loss of under-replication of pericentric heterochromatin. Multiple observations, however, indicate that we are measuring true changes in RT rather than the loss of under-replication in *Rif1*$^{-}$ follicle cells. First, loss of under-replication predicts that 100% of pericentric heterochromatin would be scored as "advanced" RT. However, we found that only 40% of pericentric heterochromatin advances RT in *Rif1*$^{-}$ mitotic and endocycling follicle cells (Figure 4.11D; Figure 4.13B). Second, if pericentric heterochromatin was under-replicated in wild type endocycling follicle cells, we would expect to observe a reduced copy number in pericentric heterochromatin relative to wild type mitotically cycling follicle cells. However, pericentric heterochromatin copy number profiles derived from wild type mitotic and endocycling S phase fractions are not different from one another (Figure 4.15). Together, these data support the conclusion that Rif1 regulates RT uniquely in different cell types and that the RT differences measured in *Rif1*$^{-}$ follicle cells represent changes in RT and do not result from changes in under-replication.

Figure 4.14. Characterization of Rif1-dependent RT control in follicle cells and wing discs.

A) Pie chart of all 100kb windows of commonly delayed RT between *Rif1*- wing discs and follicle cells. Windows located within pericentromeres are in grey and windows located within chromosome arms are in black. **B)** Pie charts of all 100kb windows with advanced (red), delayed (blue), and unchanged (grey) RT in *Rif1*$^{-}$ mitotically cycling follicle cells, endocycling follicle cells, and wing discs separated by chromosome arms (top) and pericentromeres (bottom). Experiments were performed in collaboration with Souradip Das.

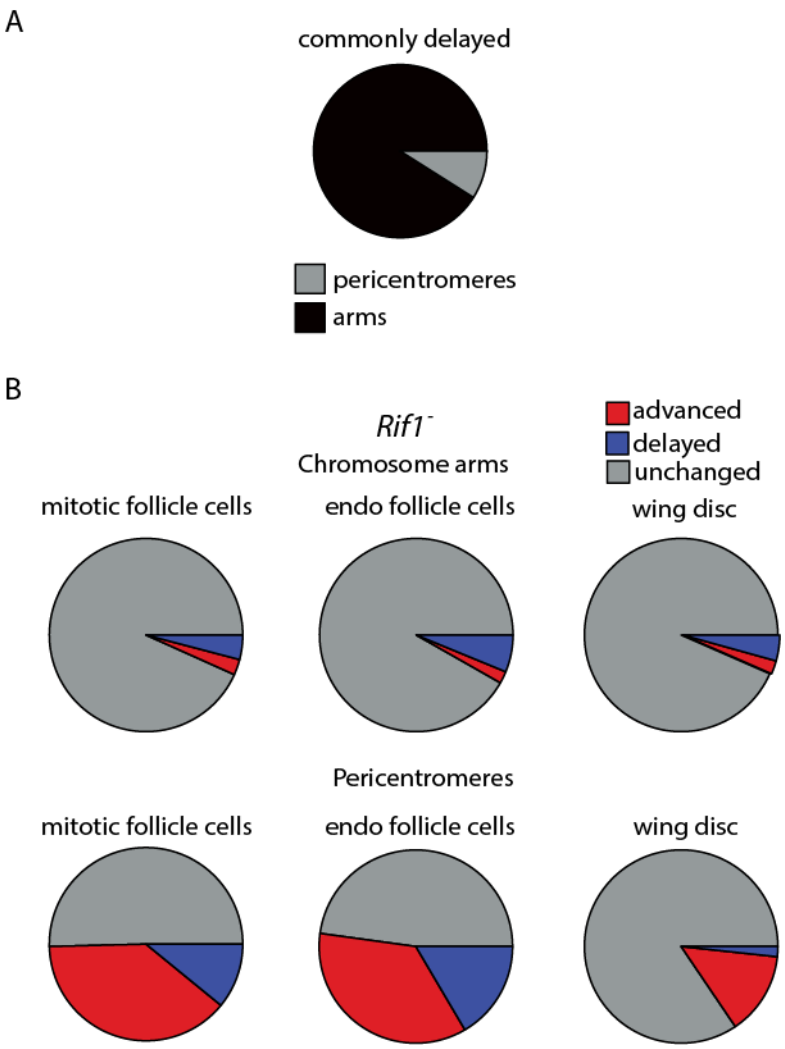
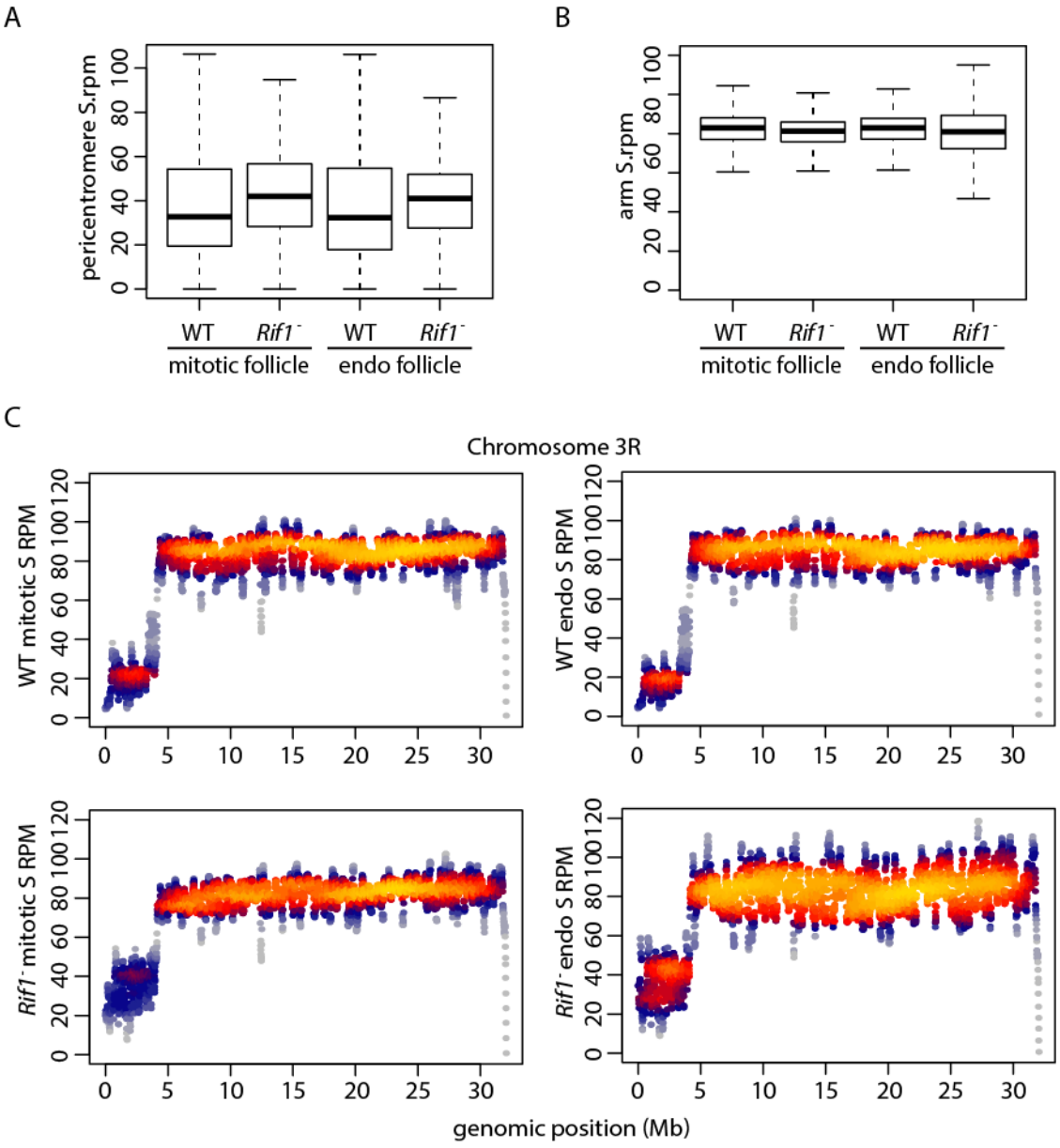

A
commonly delayed
pericentromeres
arms
B
Rif1-
Chromosome arms
advanced
delayed
unchanged
mitotic follicle cells
endo follicle cells
wing disc
Pericentromeres
mitotic follicle cells
endo follicle cells
wing disc

Figure 4.15. Under-replication does not contribute to RT differences between mitotically cycling and endocycling follicle cells.

A-B) Boxplot of S phase copy number at pericentromeres (A) and chromosome arms (B) in wild type (WT) and *Rif1*$^{-}$ mitotically cycling and endocycling follicle cells across all major chromosome scaffolds. **C)** Heatscatter plots of S phase copy number at 100kb windows with a 10kb slide across the Chromosome 3R scaffold in wild type (WT) and *Rif1*$^{-}$ mitotically cycling (left) and endocycling (right) follicle cells. Experiments were performed in collaboration with Souradip Das.


A
pericentromere S.rpm
0 20 40 60 80 100
WT
Rif1-
WT
Rif1-
mitotic follicle
endo follicle
B
arm S.rpm
0 20 40 60 80 100
WT
Rif1-
WT
Rif1-
mitotic follicle
endo follicle
C
Chromosome 3R
WT mitotic S RPM
WT endo S RPM
Rif1- mitotic S RPM
Rif1- endo S RPM
0 20 40 60 80 100 120
0 5 10 15 20 25 30
genomic position (Mb)

Rif1 controls RT independently of gene expression

To determine whether RT changes in *Rif1*$^{-}$ wing discs and follicle cells were due to transcriptional deregulation, we generated transcriptomes from *Rif1*$^{-}$ follicle cells and *Rif1*$^{-}$ wing discs. We identified only 121 and 60 differentially expressed transcripts between *Rif1*$^{-}$ and controls in wing discs and mitotic follicle cells, respectively, demonstrating that gene expression is largely unaffected after loss of Rif1 function in these tissues (Figure 4.9D, Figure 4.3). We found only 2.1% (28/1342) of differential RT windows in follicle cells and 19.5% (99/507) of differential RT windows in wing discs contain a differentially expressed transcript (Figure 4.11F). Together, these data show that loss of Rif1 function affects RT to a greater extent in follicle cells relative to wing discs and that these RT changes likely do not result from transcriptional deregulation.

Rif1's PP1 binding motif is essential for Rif1-mediated RT control

Rif1 impacts the RT of pericentric heterochromatin to a greater extent in follicle cells than in wing discs (Figure 4.11D,E) suggesting a different requirement for Rif1 in RT regulation of pericentric heterochromatin in different tissues. To further understand these mechanistic differences, we assessed what role the PP1 binding motif within Rif1 has on RT control of pericentric heterochromatin in wing discs and follicle cells. Rif1 orthologs from yeast to humans contain a PP1 binding motif, and mutation of this motif prevents Rif1 association with PP1 in multiple systems ((Davé et al. 2014; Hiraga et al. 2014; Mattarocci et al. 2014; Sreesankar et al. 2015; Alver et al. 2017; Hiraga et al. 2017; Sukackaite et al. 2017)). We previously generated an allele of Rif1 (*Rif1*^{PP1}) where the conserved SILK/RSVF PP1 interaction motif is mutated to SAAK/RASA (Munden et al. 2018). We generated genome-wide RT profiles from *Rif1*^{PP1} wing discs and follicle cells. Individual replicates

from each tissue correlated well (Pearson's correlation = 0.91 and 0.89; Figure 4.16A,B; Figure 4.17A,B). In contrast, we found that 17.9% and 11% of windows in *Rif1*^{PP1} wing discs and follicle cells, respectively, displayed differential RT relative to control (Figure 4.18A,B; Figure 4.16C,D; Figure 4.17C,D; Table 4.1). Strikingly, *Rif1*^{PP1} wing discs displayed over 3-fold the number of advanced windows compared to *Rif1*$^{-}$ wing discs. In addition, almost all (94.4%) advanced windows in *Rif1*$^{-}$ wing discs were also advanced in *Rif1*^{PP1} mutants (Figure 4.18B). Interestingly, in follicle cells, there was almost a complete overlap of advanced RT windows between *Rif1*^{PP1} and *Rif1*$^{-}$ mutants. These data suggest that the *Rif1*^{PP1} and *Rif1*$^{-}$ mutations potentially affect RT through different mechanisms in wing discs and through the same mechanism in follicle cells. In contrast, the overlap of delayed RT changes between *Rif1*^{PP1} and *Rif1*$^{-}$ wing discs or follicle cells is poor (Figure 4.18B). These data further support that advanced RT in *Rif1* mutants is a direct consequence of Rif1 loss, whereas delayed RT may be a secondary effect.

As Rif1 affects RT of pericentric heterochromatin in both tissues, we hypothesized that RT changes in *Rif1*^{PP1} tissues would preferentially be located at pericentromeres. We found that approximately 48% of pericentric heterochromatin displayed a significant advancement of RT in *Rif1*^{PP1} wing discs, unlike what we found for *Rif1*$^{-}$ null wing discs where only ~10% of pericentric heterochromatin advanced. The *Rif1*^{PP1} wing disc RT phenotype is more similar to what we observed at pericentric heterochromatin in *Rif1*$^{-}$ follicle cells (Figure 4.11A). Specifically, 80% (876/1095) of advanced windows in *Rif1*$^{-}$ mitotic follicle cells were also advanced in *Rif1*^{PP1} wing discs (Figure 4.17E). Additionally, all commonly advanced windows between *Rif1*$^{-}$ follicle cells and wing discs were advanced in *Rif1*^{PP1} wing discs. Interestingly, while the magnitude of RT change at pericentromeres is

significantly greater in $Rif1^{PP1}$ wing discs relative to $Rif1^{-}$ wing discs ($p < 2.2 \times 10^{-16}$), the magnitude of RT change in $Rif1^{PP1}$ wing discs remains significantly lower than what is observed in $Rif1^{-}$ or $Rif1^{PP1}$ follicle cells (Figure 4.18C). Collectively, these data demonstrate that the $Rif1^{PP1}$ mutation differentially affects pericentric heterochromatin RT relative to the $Rif1^{-}$ mutation in wing discs and suggest that regulatory mechanisms, potentially including the Rif1-PP1 interaction, function differently to regulate late RT of pericentromeres between tissue.

Figure 4.16. Characterization of RT in *Rif1*^{PP1} wing discs. A) Heatscatter plot comparing *Rif1*^{PP1} wing disc S/G1 (log_2) replicate replication timing values. **B)** LOESS regression line showing average S/G1 (log_2) replication timing values for wild type wing discs (black) and *Rif1*^{PP1} wing discs (gold) at 100kb windows using a 10kb slide across the major chromosome scaffolds. **C)** Pie chart of all 100kb windows of significantly advanced (red), delayed (blue), and unchanged RT (grey) in *Rif1*^{PP1} wing discs across the major chromosome scaffolds. **D)** Heatscatter plot comparing wild type and *Rif1*^{PP1} wing disc S/G1 (log_2) ratios at 100kb windows using a 10kb slide across all major chromosome scaffolds.

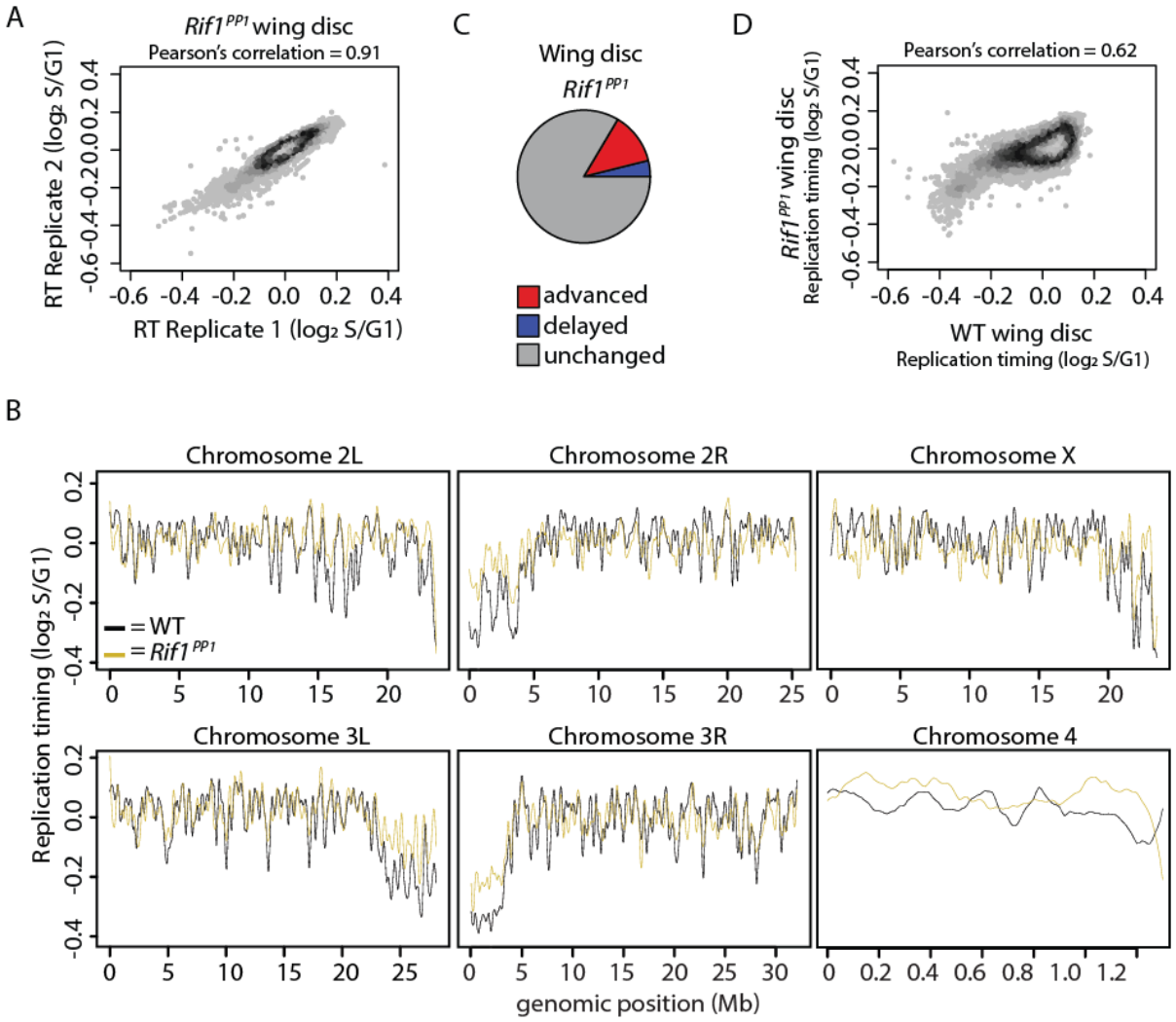


Figure 4.17. Characterization of RT in *Rif1*^{PP1} mitotic follicle cells. **A)** Heatscatter plot comparing *Rif1*^{PP1} mitotic follicle cell S/G1 (log_2) replicate replication timing values. **B)** LOESS regression line showing average S/G1 (log_2) replication timing values for wild type mitotic follicle cells (black) and *Rif1*^{PP1} mitotic follicle cells (gold) at 100kb windows using a 10kb slide across the major chromosome scaffolds. **C)** Pie chart of all 100kb windows of significantly advanced (red), delayed (blue), and unchanged RT (grey) in *Rif1*^{PP1} mitotic follicle cells across the major chromosome scaffolds. **D)** Heatscatter plot comparing wild type and *Rif1*^{PP1} mitotic follicle cell S/G1 (log_2) ratios at 100kb windows using a 10kb slide across all major chromosome scaffolds. **E)** Venn diagram comparing advanced 100kb windows between *Rif1*^{PP1} wing discs (gold) and *Rif1*$^{-}$ mitotic follicle cells (cyan). Experiments were performed in collaboration with Souradip Das.

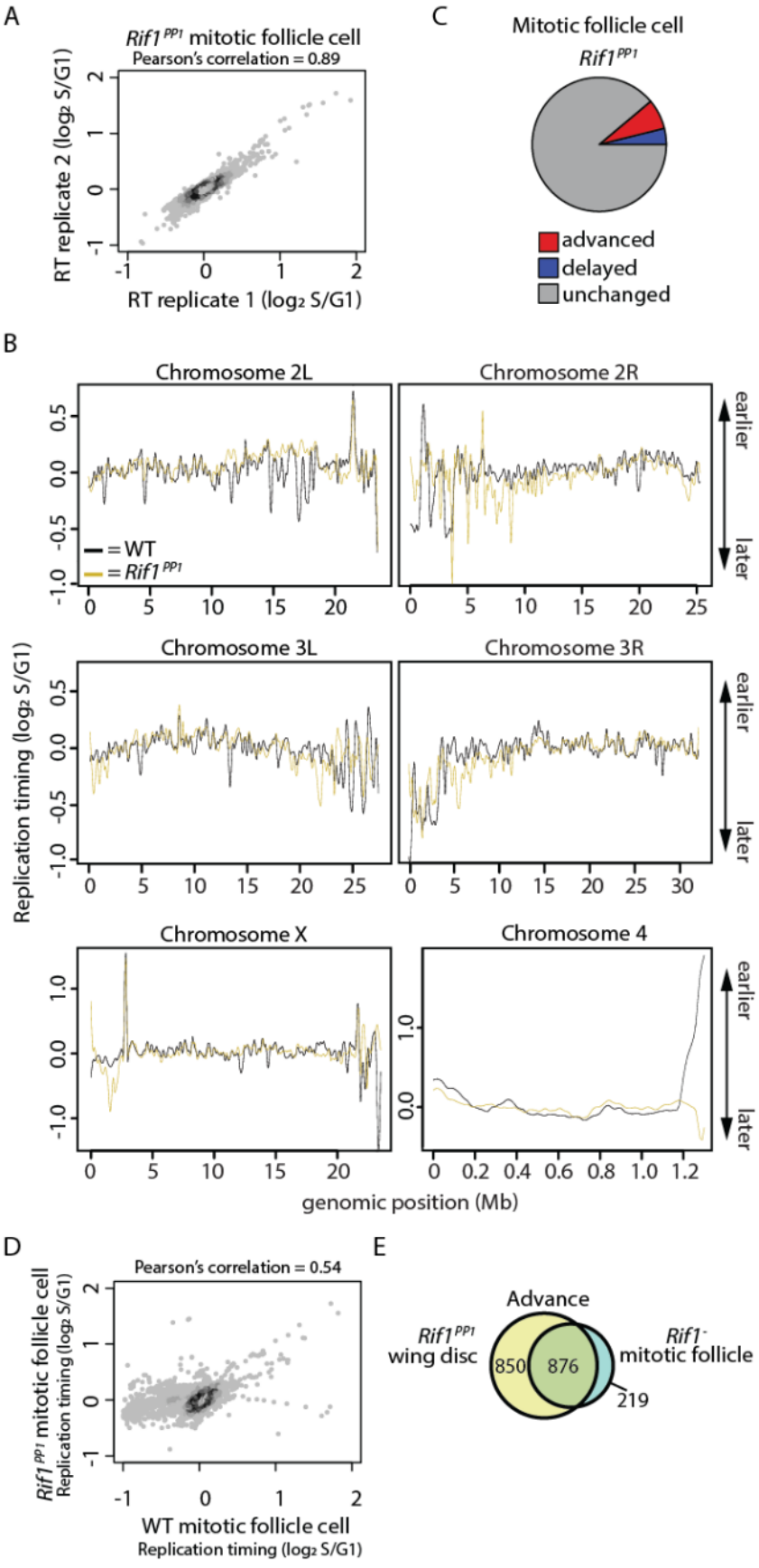
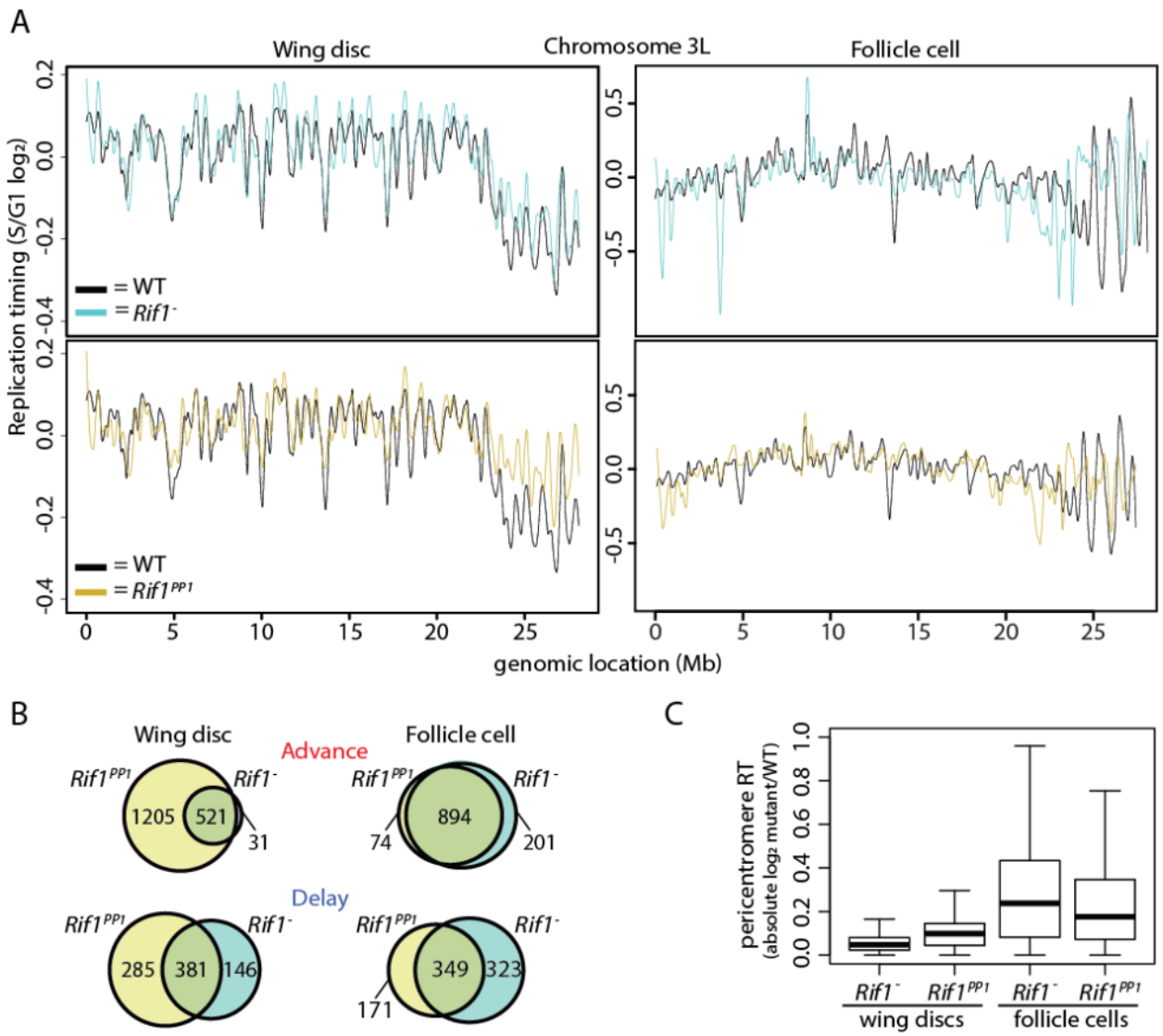

A
Rif1 PP1 mitotic follicle cell
Pearson's correlation = 0.89
RT replicate 2 (log2 S/G1)
RT replicate 1 (log2 S/G1)
C
Mitotic follicle cell
Rif1 PP1
advanced
delayed
unchanged
B
Chromosome 2L
Chromosome 2R
Chromosome 3L
Chromosome 3R
Chromosome X
Chromosome 4
= WT
= Rif1 PP1
earlier
later
Replication timing (log2 S/G1)
genomic position (Mb)
D
Pearson's correlation = 0.54
Rif1 PP1 mitotic follicle cell
Replication timing (log2 S/G1)
WT mitotic follicle cell
Replication timing (log2 S/G1)
E
Advance
Rif1 PP1
wing disc
850
876
Rif1-
mitotic follicle
219

Figure 4.18. Rif1's PP1 binding motif is essential for Rif1-mediated RT control. A) LOESS regression line showing average *Rif1*$^{-}$ (cyan), *Rif1*^{PP1} (gold), and wild type (black) S/G1 (log_2) replication timing values in wing discs (left) and follicle cells (right) across the chromosome 3L scaffold. See Figures S5, S6, S11, and S12 for other chromosomes. **B)** Venn diagrams comparing significantly advanced (top) and delayed (bottom) 100kb windows identified in *Rif1*$^{-}$ (cyan) and *Rif1*^{PP1} (gold) wing discs (left) and follicle cells (right) ($p<0.01$ and absolute log_2 fold change > 0.1; limma). **C)** Box plot of absolute mutant/control log_2 ratio of normalized replication timing values (S/G1 (log_2)) at all pericentromeric regions of the major chromosome scaffolds. Experiments were performed in collaboration with Souradip Das.


A
Wing disc
Chromosome 3L
Follicle cell
Replication timing (S/G1 log2)
= WT
= Rif1-
= WT
= Rif1PP1
genomic location (Mb)
B
Wing disc
Follicle cell
Advance
Rif1PP1
Rif1-
1205
521
31
894
74
201
Delay
285
381
146
349
323
171
C
pericentromere RT
(absolute log2 mutant/WT)
Rif1-
Rif1PP1
Rif1-
Rif1PP1
wing discs
follicle cells

Discussion

Our findings provide insight into the relative contributions that cell type, gene expression, cell cycle, and Rif1 make to RT control. By comparing genome-wide RT profiles from unperturbed cells from distinct tissues, we demonstrated that cell type has a larger effect on RT than loss of Rif1, an evolutionarily conserved regulator of RT. We also found that the RT program is not modified in response to the physiological and transcriptional changes that occur during the mitotic-to-endocycle transition.

We found that ~30% of the genome had different RT in the two tissue types we examined, and that transcriptional changes do not account for these changes. Studies in other systems also have failed to establish a direct relationship between changes in RT and changes in transcriptional activity (MacAlpine et al. 2004; Lubelsky et al. 2014; Siefert et al. 2017; Almeida et al. 2018; Armstrong et al. 2018). While transcriptional activity has long been correlated with RT, there are clearly mechanisms that control RT independently of transcription. RT is highly correlated with genome topology (Pope et al. 2014), and recent work has demonstrated that changes in TAD structure can be uncoupled from changes in gene expression (Ghavi-Helm et al. 2019). Therefore, our results are consistent with a model in which lineage-specific changes in genome topology, not transcription, underlie changes to the RT program as cells differentiate. These RT programs can then further be enforced by *trans*-acting factors such as Rif1.

When comparing different tissues, we found a higher degree of overlap between regions of the genome that transition from late-to-early in the absence of Rif1 than those that transition from early-to-late (Figure 4.11A). These data imply that Rif1 directly promotes late replication of specific regions of the genome while indirectly affecting regions of the genome

that normally replicate early. Rif1 dynamically associates with heterochromatin from yeast to humans (Buonomo et al. 2009; Seller et al. 2019). In early *Drosophila* embryos, Rif1 is recruited to heterochromatic regions independently of HP1a, and then displaced from heterochromatin immediately before heterochromatin is replicated late in S phase (Seller and O'Farrell 2018). Chromatin immunoprecipitation of Rif1 followed by sequencing have revealed that in yeast and mouse cells Rif1 targets many other regions of the genome with both late and early replicating domains (Hayano et al. 2012; Foti et al. 2016). It is currently unknown, however, how Rif1 is targeted to heterochromatin and other late-replicating regions of the genome to delay RT. Our results argue that Rif1 localization to chromatin is likely influenced by cell type-specific factors.

Our results demonstrate that in metazoans the PP1 interaction motif of Rif1 can contribute to Rif1-mediated RT control. These data suggest that helicase inactivation, or inactivation of another PP1 target near origins of replication, is critical for Rif1-mediated RT control. Multiple models have been proposed to explain how Rif1 controls RT. First, through a direct interaction with PP1, Rif1 is thought to counteract DDK-mediated helicase activation and delay replication of Rif1-associated regions. Second, based on 4C experiments with five viewpoints, Rif1 was shown to affect chromatin contacts between different RT domains, suggesting that Rif1 controls RT through nuclear organization. It is unclear how these different models are related, if at all. Furthermore, while the timing decision point occurs in G1 phase, helicase activation occurs throughout S phase, raising additional mechanistic questions about how Rif1 controls RT. Recent work in budding yeast has shown that DDK can act in G1 phase. Additionally, DDK-dependent helicase activation, and Cdc45 recruitment, in G1 phase is critical for the specification of certain replication origins. Thus,

premature helicase activation in the absence of Rif1 during G1 phase could alter the localization of specific replication domains. While this model could unify the observations describing how Rif1 controls RT, further work is needed to test this possibility.

Our data suggest that different regulatory mechanisms control late RT between wing discs and follicle cells. The approximately 3-fold increase in the number of windows with advanced RT in *Rif1*^{PP1} wing discs relative to *Rif1*$^{-}$ null wing discs was surprising. These data indicate that the presence of mutant Rif1^{PP1} protein results in a stronger effect that the absence of Rif1. One possibility is that Rif1^{PP1} acts in a dominant negative manner in regions of the genome that normally replicate late during S phase, such as pericentric heterochromatin. Another striking observation was that loss of Rif1 in wing discs did not substantially advance RT in much of the pericentric heterochromatin. This result suggests that mechanisms in addition to Rif1/PP1-mediated MCM dephosphorylation act within the wing disc to promote late replication of pericentric heterochromatin.

In summary, our study demonstrates that cell lineage is a major driver of RT control within the context of a developing organism. Rif1 fine tunes the RT program established in different cell types, and each of these modes of RT control function independently of transcriptional control, suggesting additional levels of regulation.

Acknowledgements

This work was supported by NIH Grants R01-GM124201 to R.J.D and NSF MCB 1818019 to J.T.N. In addition, R.L.A. was supported in part by an NIH predoctoral training grant T32-GM007092. We thank the UNC Flow Cytometry and High Throughput Sequencing Core Facilities, supported in part by P30 CA016086 Cancer Center Core Support

Grant to the UNC Lineberger Comprehensive Cancer Center. FACS results reported in this publication were supported in part by the North Carolina Biotechnology Center Institutional Support Grant 2012-IDG-1006. Flow Cytometry experiments were performed in the VMC Flow Cytometry Shared Resource. The VMC Flow Cytometry Shared Resource is supported by the Vanderbilt Ingram Cancer Center (P30 CA68485) and the Vanderbilt Digestive Disease Research Center (DK058404).

CHAPTER 5- EXPLORING DISTINCT ROLES FOR H3K9 AND RIF1 IN PERICENTRIC HETEROCHROMATIN DNA REPLICATION

Introduction

To accurately and completely duplicate eukaryotic genomes each cell division cycle, replication initiates from many discrete loci throughout the genome, termed replication domains. Genome-wide initiation of DNA replication does not occur synchronously in space and time, yielding a temporal pattern of early and late replicating domains known as the DNA replication timing (RT) program. Accessible, transcriptionally active euchromatin tends to replicate early during S phase, whereas inaccessible, transcriptionally repressive heterochromatin tends to replicate late during S phase (Lubelsky et al. 2014). Chromatin states are largely defined by signatures of histone post-translational modifications (PTMs) such that euchromatin is generally enriched in H3 and H4 acetylation (ac) and heterochromatin is enriched in di- and tri-methylation of lysine nine of histone H3 (H3K9me2/3) and tri-methylation of lysine twenty-seven of histone H3 (H3K27me3) (Kharchenko et al. 2011). Additionally, many non-histone proteins contribute to the formation and function of different chromatin states. For example, recent work in mammalian cells identified approximately 170 proteins enriched in H3K9me3-marked heterochromatin (Becker et al. 2017). However, the relative contributions of histone and non-histone heterochromatin-associated factors to RT control remains poorly understood.

Rif1 is a conserved RT control factor across eukaryotic species that promotes late RT of heterochromatin (Cornacchia et al. 2012; Hayano et al. 2012; Yamazaki et al. 2012; Seller

and O'Farrell 2018). To promote late RT, Rif1 associates with heterochromatin and recruits Protein Phosphatase 1 (PP1) that de-phosphorylates the MCM replicative helicase during early S phase which prevents DNA replication initiation in heterochromatic domains in early S phase (Davé et al. 2014; Hiraga et al. 2014; Mattarocci et al. 2014; Hiraga et al. 2017; Sukackaite et al. 2017). In late S phase, Rif1 and PP1 dissociate from chromatin via CDK-mediated phosphorylation of Rif1, allowing replication of heterochromatic domains at the end of S phase (Seller and O'Farrell 2018). Further, Rif1 has been shown to contribute to the establishment of a three-dimensional nuclear architecture that promotes proper RT in mammalian systems (Foti et al. 2016). Importantly, Rif1 is required for the onset of late replication in early embryogenesis in *Drosophila* and precedes, but is not required for, the establishment of H3K9me/HP1a enriched constitutive heterochromatin (Seller and O'Farrell 2018). These data suggest that Rif1 may regulate pericentric heterochromatin RT independently from H3K9.

We previously took a genetic approach to understand the relative contribution of heterochromatin-associated RT regulators, including chromatin structure and the eukaryotic RT control factor, Rif1. To perturb local chromatin structure, we utilized our genetic platform in which endogenous histone genes are replaced with transgenic histone genes encoding mutations that prevent modification of specific histone residues (McKay et al. 2015; Armstrong et al. 2018). Despite well-established correlations between early replication and euchromatin, and late replication and heterochromatin, we observed that perturbation of heterochromatin through *H3K9R* mutation does not result in large-scale changes in RT, suggesting critical regulation beyond H3K9 (Armstrong et al. 2018). Furthermore, we profiled RT in the absence of Rif1 and found that while loss of Rif1 affected genome-wide

RT to a greater extent than perturbation of chromatin structure, pericentric heterochromatin remained late replicating relative to the mostly euchromatic chromosome arms. These data demonstrate that loss of H3K9 or Rif1 alone is insufficient to perturb late replication of pericentric heterochromatin.

Here, we investigate the relationship between H3K9 and Rif1 in regulating late replication of pericentric heterochromatin using cytological and genomic approaches in *Drosophila*. We find that both H3K9 and Rif1 are required for S phase progression. Furthermore, we identify that H3K9 and Rif1 regulate RT of distinct heterochromatic replication domains and appear to affect heterochromatin replication, at least partially, through distinct mechanisms. Future study is necessary to understand the extent of functional redundancy between H3K9 and Rif1 in heterochromatin replication.

Materials and Methods

Immunofluorescence

Third instar wandering larvae were dissected and the carcasses inverted to expose attached imaginal discs. Tissues were incubated for 60' in 0.1mg/mL EdU. Tissues were then fixed in 3.7% paraformaldehyde in PBS for 25 min. EdU incorporation was detected using the Click-It EdU Alexa Fluor 488 Imaging Kit (ThermoFisher Scientific). DNA was stained with DAPI, and the discs were mounted in ProLong Gold antifade reagent and imaged on a Leica confocal microscope.

FAIRE

FAIRE-seq samples from female third instar wing imaginal discs were prepared as described previously (McKay and Lieb 2013; Penke et al. 2018). Libraries were prepared

with the Rubicon ThruPLEX DNA-seq kit. Sequencing was performed on an Illumina HiSeq 2500.

Sequence Data Analysis

RT data under the Gene Expression Omnibus (GEO) accession number GSE141632 were used for comparisons in this study. FAIRE-seq samples were aligned to the dm6 reference genome (release 6.04) using Bowtie2 default parameters and filtered by a MAPQ score ≥ 10 (Langmead et al. 2009). FAIRE peaks were called using MACS2 and FAIRE-seq signal at peaks was normalized to sequencing depth (Zhang et al. 2008). Differential signal analysis on peaks was performed using edgeR (Robinson et al. 2010). For genotype comparisons, only peaks called in both replicates were used for further analysis.

Results

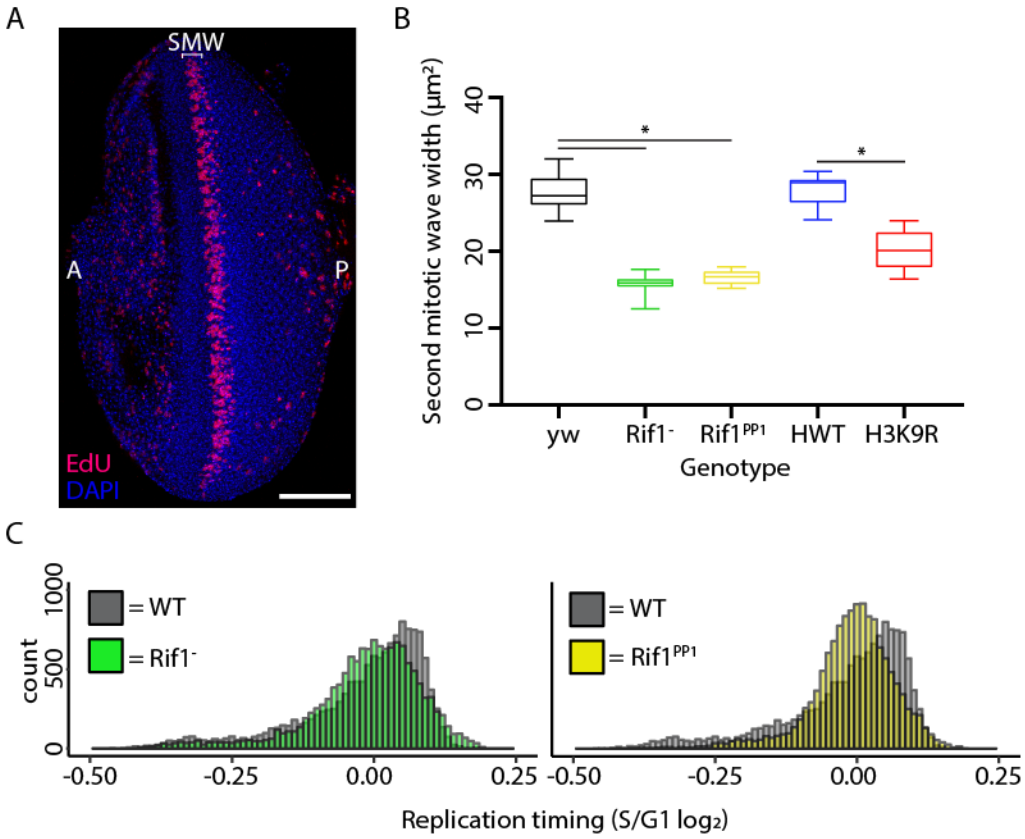
H3K9 and Rif1 regulate S phase progression

We first asked whether H3K9 and Rif1 regulate S phase progression in *Drosophila*. Previous studies have shown that mutation of the H3K9me2/3 writer (Su(var)3-9) or of a H3K9me2/3 reader (Heterochromatin Protein 1a (HP1a)) results in cell cycle defects, including a decreased S-phase index and an increased G2/M index (De Lucia et al. 2005; Pindyurin et al. 2008; Quivy et al. 2008; Peng and Karpen 2009; Schwaiger et al. 2010; Sidler et al. 2014). In agreement with these data, we have previously shown that *Drosophila H3K9R* mutants display a modest cell cycle defect, with a decrease in the proportion of S phase cells and a concomitant increase in the proportion of G2/M phase cells. These data show that H3K9 directly promotes cell cycle progression (Armstrong et al. 2018). We were curious as to whether Rif1, a heterochromatin-associated RT control factor, also regulates S

phase progression. To address this question, we took advantage of a unique cell cycle strategy in the eye imaginal disc.

During larval development, a wave of differentiation migrates posterior to anterior across the eye disc, termed the morphogenetic furrow. In the morphogenetic furrow, cells arrest in G1 and a subset begin differentiating; upon exit from the morphogenetic furrow, the remaining undifferentiated cells enter a synchronous cell division cycle termed the second mitotic wave (SMW) (Figure 5.1A) (Ready et al. 1976; Sukhanova and Du 2008). As migration of the morphogenetic furrow occurs at a defined rate, the relative duration of S phase within the SMW can be determined by measuring the width of the SMW following a 60-minute pulse of EdU. We performed eye disc SMW analysis and, similar to our previous flow cytometry analysis in *H3K9R* mutants, found that *H3K9R* mutants have a decreased SMW width relative to controls (Figure 5.1B). From these data we conclude that *H3K9R* mutants have a decreased S phase duration. We next performed EdU pulsing and subsequent SMW measurement in $Rif1^{-}$ (a null allele of *Rif1*) and $Rif1^{PP1}$ (an allele of *Rif1* that prevents the interaction between Rif1 and PP1) mutants previously generated by the Nordman lab (Munden et al. 2018). We found that both $Rif1^{-}$ and $Rif1^{PP1}$ mutants have a shorter SMW width relative to controls (Figure 5.1B). These data suggest that factors that regulate heterochromatin RT may also promote proper S phase progression. We hypothesized that the shortened S phase duration observed in *Rif1* and *H3K9R* mutants may be due to defects in heterochromatin RT. However, whereas $Rif1^{-}$ and *H3K9R* wing discs display far fewer RT changes within pericentric heterochromatin relative to $Rif1^{PP1}$ wing discs (Armstrong et al. 2018) (Chapter 4), all three mutants similarly affect S phase progression suggesting that RT is not the major contributor to the observed shortening of S phase duration.

Figure 5.1. Rif1 and H3K9 promote S phase progression. A) Representative eye imaginal disc stained for EdU (magenta) and counterstained with DAPI (blue) with the anterior (A) and posterior (P) axes indicated. White bracket indicates the location of the second mitotic wave (SMW). The scale bar is set to 50µM. **B)** Quantification of the average width of the SMW for *yw* (wild type; black), *Rif1*$^{-}$ (green), *Rif1*^{PP1} (yellow), *HWT* (blue), and *H3K9R* (red). For each biological replicate, ten measurements taken along the length of the SMW of one eye disc were averaged. **C)** Histogram of RT values for wild type (grey), *Rif1*$^{-}$ (green), and *Rif1*^{PP1} (yellow).

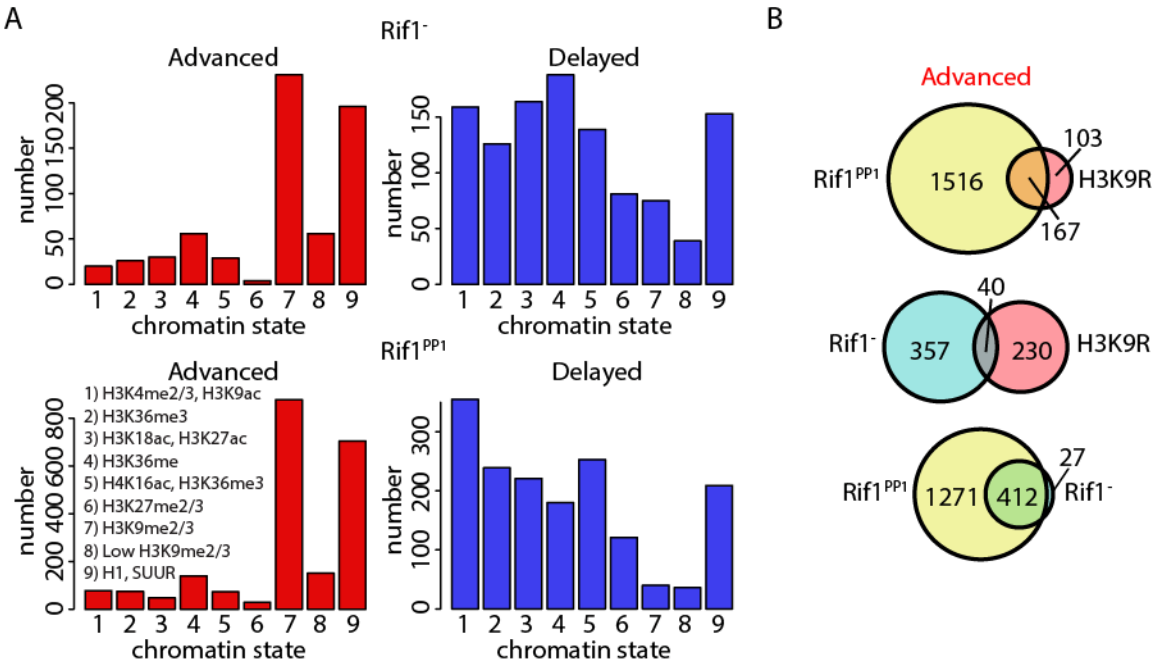


H3K9 and Rif1 regulate RT of independent heterochromatic domains

We were curious as to whether Rif1 and H3K9 regulate replication through redundant mechanisms. To address this, we first directly compared the genome-wide RT distribution in *Rif1*$^{-}$ and *H3K9R* mutants. While the distribution of RT values for *H3K9R* mutants is similar to controls, we were surprised to find that the RT program in both *Rif1*$^{-}$ and *Rif1*^{PP1} mutants is condensed relative to controls (Figure 5.1C). These data suggest that Rif1 may regulate RT distinctly from H3K9. We next asked if Rif1 and H3K9 regulate RT of the same genomic regions. Assigning RT changes identified in *H3K9R* and *Rif1* mutants to one of nine previously described chromatin states revealed that, while almost all advanced RT changes in *H3K9R* mutants occur in H3K9me2/3-enriched heterochromatin (Armstrong et al. 2018), advanced RT changes in *Rif1*$^{-}$ mutants primarily occur either in H3K9me2/3-enriched heterochromatin or H1/SuUR-enriched "black" heterochromatin, suggesting that Rif1 and H3K9 affect RT of different heterochromatic regions (Figure 5.2A) (Kharchenko et al. 2011).

To test this hypothesis, we directly compared advanced RT changes identified in *H3K9R* and Rif1 mutants. While *Rif1*$^{-}$ and *H3K9R* mutants disrupt RT at a relatively similar number of windows across the genome, the advanced RT changes in *Rif1*$^{-}$ and *H3K9R* mutants are largely non-overlapping (Figure 5.2B). In contrast, approximately 62% of advanced RT changes observed in *H3K9R* mutants are also advanced in *Rif1*^{PP1} mutants, most likely because 48% of all pericentric heterochromatin advances its RT in *Rif1*^{PP1} wing discs (Figure 5.2B). Together, these data demonstrate that Rif1 and H3K9 regulate replication of distinct heterochromatin domains. However, the extent of redundancy in regulating pericentric RT between Rif1 and H3K9 remains untested.

Figure 5.2. Rif1 and H3K9 regulate RT of unique heterochromatic domains. A) All advanced (red) 10kb windows in *Rif1*$^{-}$ and *Rif1*^{PP1} mutants were assigned to the nine chromatin states defined in flies (Kharchenko et al. 2011). Shown are the number of windows that overlap each chromatin state. **B)** Venn diagram comparisons of significantly advanced 10kb windows identified in *H3K9R* (red), *Rif1*$^{-}$ (cyan), and *Rif1*^{PP1} (yellow) mutant wing imaginal discs ($p < 0.01$, absolute $\log_2$ fold change > 0.1; limma) (Armstrong et al. 2018) (Chapter 4).



H3K9 and Rif1 differentially affect chromatin accessibility

Chromatin accessibility and transcriptional activity are thought to be key regulators of replication initiation. To test this model, we previously explored the extent to which transcription and chromatin accessibility contribute to RT using *H3K9R* mutants. We found that while almost all advanced RT changes in *H3K9R* mutants occurred at domains of increased chromatin accessibility and transcriptional activity, only a very small percentage of the chromatin accessibility and transcriptional changes experienced a coincident RT change (Armstrong et al. 2018). These data demonstrate that chromatin accessibility and transcriptional activity are insufficient for RT change, and, rather, are permissive regulators of replication initiation. Because chromatin accessibility and transcriptional activity contribute to RT change in *H3K9R* mutants, we were curious as to whether these same biological processes contribute to RT change in *Rif1*$^{-}$ mutants.

We previously identified that RT change in *Rif1*$^{-}$ mutants occurred independently from transcriptional change as *Rif1*$^{-}$ mutants experience minimal transcriptional deregulation (Chapter 4). Therefore, we next profiled chromatin accessibility by performing Formaldehyde-Assisted Isolation of Regulatory Elements followed by sequencing (FAIRE-seq) from *Rif1*$^{-}$ and wild type control wing imaginal discs in duplicate. We confirmed the genotypes within the wild type and *Rif1*$^{-}$ FAIRE-seq datasets and observed high correlation between individual replicates for both wild type (Pearson's correlation = 0.90) and *Rif1*$^{-}$ (Pearson's correlation = 0.97) mutants (Figure 5.3A,B). We identified 32,648 total FAIRE peaks in wild type and 29,354 total FAIRE peaks in *Rif1*$^{-}$ that were shared between replicates. Direct comparison of wild type and *Rif1*$^{-}$ FAIRE peaks revealed that 91.9% of peaks identified in *Rif1*$^{-}$ wing discs were also identified in wild type suggesting that open

chromatin profiles are highly similar between *Rif1*$^{-}$ and control (Figure 5.3C). We next directly compared individual FAIRE peaks between genotypes to identify differential peak intensities between *Rif1*$^{-}$ and control. Our edgeR analysis revealed 80 significantly different peaks between *Rif1*$^{-}$ and wild type control (Figure 5.3D). Of those 80 FAIRE peaks, 30 were more "open" and 50 were more "closed" in *Rif1*$^{-}$ mutants. Interestingly, 99.9% of FAIRE peaks unique to either wild type or *Rif1*$^{-}$ wing discs and 90% of differential FAIRE peaks were located along the chromosome arms demonstrating that the small amount of differential chromatin accessibility is localized to the mostly euchromatic portion of the genome.

To ascertain whether RT changes in *Rif1*$^{-}$ mutants were occurring at either sites of *de novo* chromatin accessibility or sites of chromatin accessibility loss, we compared FAIRE peaks unique to *Rif1*$^{-}$ mutants or WT, respectively, to 10kb windows of RT change. We observed a slight enrichment of *de novo* FAIRE peaks at windows of RT change in *Rif1*$^{-}$ mutants as 27.9% of 10kb windows with an RT change contain a *de novo* FAIRE peak, relative to 15.8% of a shuffle set of 10kb windows (Figure 5.4). Additionally, we observed an enrichment of chromatin accessibility loss at windows of RT change, with 29.6% of significant RT changes containing a FAIRE peak identified in wild type that is lost in *Rif1*$^{-}$ mutants, relative to a 9.1% of a shuffle set of windows (Figure 5.4). These data suggest that a *de novo* gain or a loss of chromatin accessibility influences RT in *Rif1*$^{-}$ mutants.

Figure 5.3. Chromatin accessibility profiling in *Rif1*$^{-}$ mutants. A) Genome browser of FAIRE-seq signal at the Rif1 locus. The red box indicates the coordinates of the Rif1 deletion allele used. **B)** Heatscatter plot comparing wild type FAIRE-seq replicates (left) and *Rif1*$^{-}$ FAITRE-seq replicates (right). **C)** Venn diagram comparing FAIRE-seq peaks between wild type and *Rif1*$^{-}$. Peaks included for each genotype were shared between replicates. **D)** Heatscatter plot of the *Rif1*$^{-}$/control ratio of total FAIRE-seq signal at all peaks. Statistically different FAIRE peaks are indicated in red ($p < 0.01$; edgeR). Blue lines indicate a $\log_2$ fold change of 1 and -1.

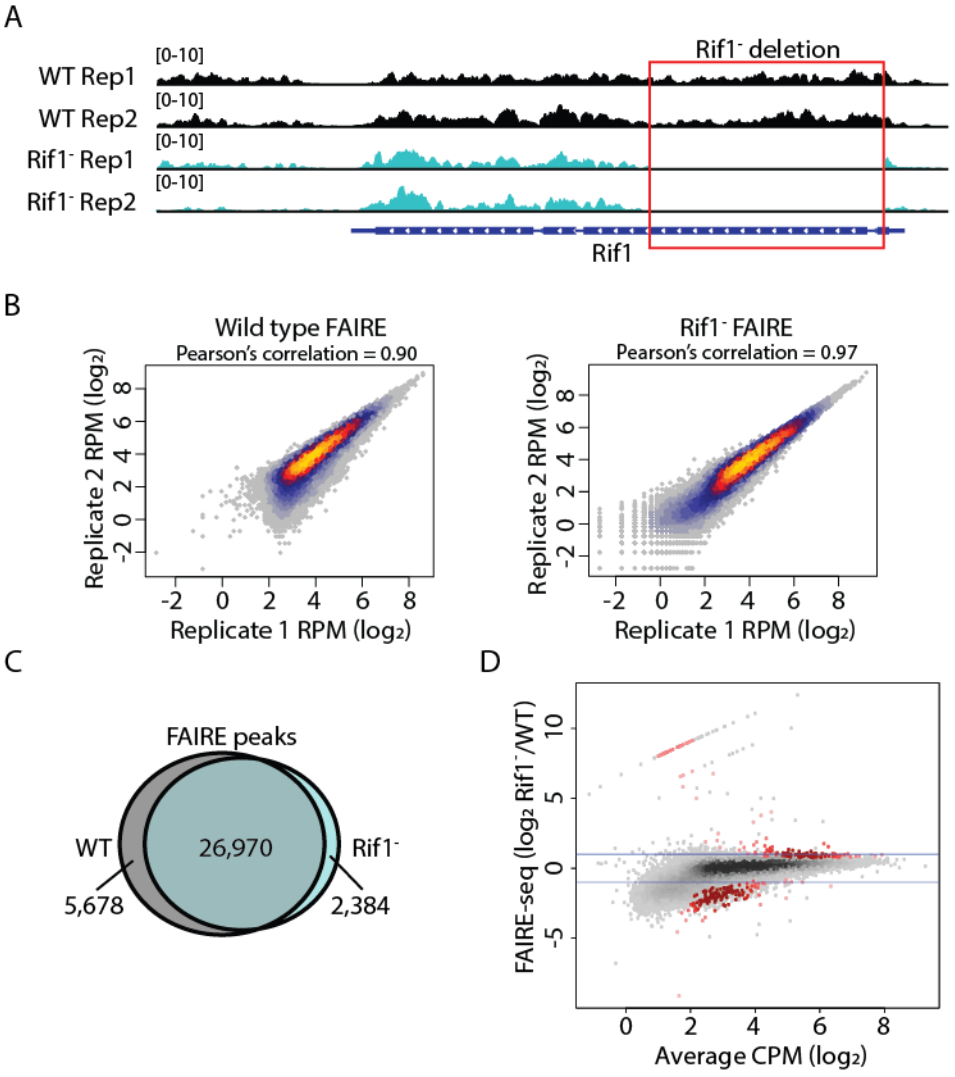
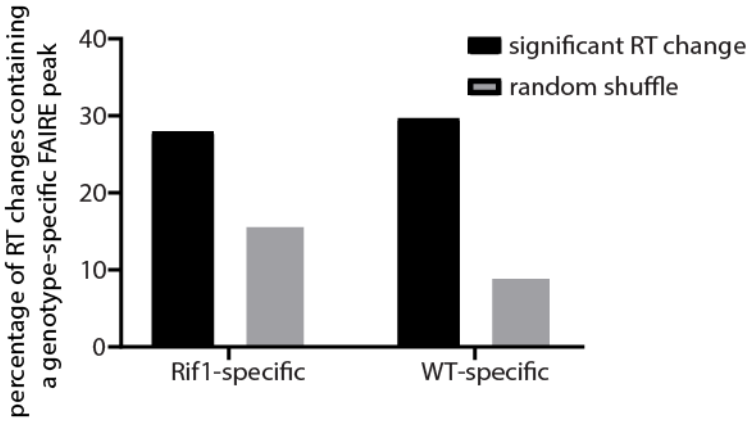

A
Rif1- deletion
WT Rep1
WT Rep2
Rif1- Rep1
Rif1- Rep2
[0-10]
Rif1
B
Wild type FAIRE
Pearson's correlation = 0.90
Rif1- FAIRE
Pearson's correlation = 0.97
Replicate 2 RPM (log2)
Replicate 1 RPM (log2)
C
FAIRE peaks
WT
Rif1-
26,970
5,678
2,384
D
FAIRE-seq (log2 Rif1-/WT)
Average CPM (log2)

Figure 5.4. A modest relationship exists between chromatin accessibility and RT in *Rif1*$^-$ mutants.

Bar plot of the percentage of 10kb windows with significantly altered RT in *Rif1*$^-$ mutants containing a FAIRE peak unique to *Rif1*$^-$ mutants (left; "opened" in *Rif1*$^-$ mutants) or unique to WT (right; "closed" in *Rif1*$^-$ mutants) ($p < 0.01$, absolute $\log_2$ fold change > 0.1; limma).



Discussion

While constitutive heterochromatin is required for maintenance of genome integrity, it is quite dynamic in nature, disassembling each mitosis and reassembling during interphase each cell division cycle. Genome function must be fully restored by the onset of interphase to ensure that genome integrity is maintained. One such process that occurs at the onset of G1 is the establishment of the genome-wide RT program, termed the timing decision point (Dimitrova and Gilbert 1999). It is therefore imperative that distinct mechanisms orchestrate the reassembly of heterochromatin in a timely manner following mitosis. Because many proteins function within constitutive heterochromatin, the relative contributions of each factor to establishment of both heterochromatin and the genome-wide RT program remain poorly understood. We therefore sought out to uncover the relative contributions of two heterochromatin-associated factors, H3K9 and Rif1, to chromatin accessibility, transcription, and RT in *Drosophila*.

Here, we provide evidence that H3K9 and Rif1 regulate heterochromatin structure, and RT control, through distinct mechanisms. We previously identified that mutation of *H3K9* results in deregulation of transposon expression and chromatin accessibility specifically at pericentric heterochromatin (Penke et al. 2016). Furthermore, almost all advanced RT changes observed in *H3K9R* mutants occur in pericentric heterochromatin specifically at domains of increased nucleosome accessibility and elevated transposon expression (Armstrong et al. 2018). In contrast, mutation of *Rif1* results in deregulation of RT at approximately 8% of the genome without largely affecting transcription or chromatin accessibility genome-wide (Chapter 4; Figure 5.3). Importantly, RT of domains within pericentric heterochromatin regulated by H3K9 and Rif1 are largely non-overlapping,

supporting a model in which H3K9 and Rif1 regulate RT of pericentric heterochromatin through distinct mechanisms.

We are not the first to suggest that H3K9 and Rif1 function independently of one another in regard to heterochromatin structure. Previous work has demonstrated that, in the absence of Rif1, HP1a properly localizes to pericentric heterochromatin in *Drosophila* suggesting that the "hallmark" structural units of heterochromatin remain intact (Sreesankar et al. 2015; Munden et al. 2018; Seller and O'Farrell 2018). Our data here further support that heterochromatin structure remains intact in the absence of Rif1, as a marginal proportion of the chromatin accessibility changes in *Rif1*$^{-}$ mutants occurred within pericentric heterochromatin. However, Rif1 has been shown to promote proper three-dimensional nuclear architecture at the timing decision point in G1, which has been proposed to directly regulate the RT program in S phase (Foti et al. 2016). It therefore is possible that Rif1 is orchestrating its RT control at the level of the 3D genome as opposed to at the level of local chromatin structure as observed in *H3K9R* mutants. It remains unknown how either Rif1 or H3K9 regulate the three-dimensional genome structure in *Drosophila*.

HP1a, a "hallmark" of constitutive heterochromatin, has been show to undergo liquid-like phase separation in both *Drosophila* and humans (Larson et al. 2017; Strom et al. 2017). We previously demonstrated that, in *H3K9R* mutants, HP1a fails to concentrate at pericentric heterochromatin (Penke et al. 2016). Interestingly, in *H3K9R* mutants, we still observed DAPI-bright heterochromatic foci suggesting that phase separation of HP1a is not required for condensation of DNA in *Drosophila* (Penke et al. 2016; Penke et al. 2018). However, because transcription and chromatin accessibility are disrupted in *H3K9R* mutants, we propose that certain heterochromatic features regulate different aspects of heterochromatin

biology. This model predicts that perturbation of multiple heterochromatin factors concurrently may disrupt heterochromatin biology to a greater extent than any one factor alone. It would be interesting to determine the effect on heterochromatin structure and function within the context of concurrent H3K9/HP1a and Rif1 loss.

The complex nature of heterochromatin renders it difficult to understand the mechanisms regulating RT within this region of the genome. To completely understand replication of constitutive heterochromatin and how RT control mechanisms may differ between euchromatic and heterochromatic regions of the genome, future studies must identify and map the molecular events dictating replication origin licensing and activation within heterochromatic domains. Furthermore, constitutive heterochromatin displays an enrichment of hundreds of proteins relative to euchromatic regions of the genome, but the extent to which these factors serve heterochromatin-specific roles is largely unknown. Our work here, identifying contributions of Rif1 and H3K9 to heterochromatin replication, begins to reveal answers regarding how independent pathways within heterochromatin contribute to its RT.

CHAPTER 6- DISCUSSION AND FUTURE DIRECTIONS

In April 2003, the human genome sequence was finalized, a mere fifty-five years after Matthew Meselson and Franklin Stahl demonstrated the semi-conservative nature of DNA replication (Watson and Crick 1953; Meselson and Stahl 1958; Lander et al. 2001; Venter et al. 2001). In just the last seventeen years, the era of whole-genome sequencing and the evolution of sequencing-based technologies have provided unprecedented advances for the fields of epigenetics and DNA replication. Hopefully this project and the future directions that this work inspires will contribute to the technological and mechanistic advances to come. The following chapter will summarize the work included in this dissertation and a selection of future directions that I propose as a follow-up to my PhD thesis work.

This work sought to discern correlative versus causative relationships between chromatin structure, transcriptional activity, and RT using genetic and genomic approaches. We identified that chromatin accessibility and transcriptional activity are mechanistically separable from replication initiation using two independent genetic strategies: 1) mutation of histone residues and 2) mutation of the RT control factor, Rif1. In the context of *H3K9R* mutation, we observe a global increase in chromatin accessibility and transposon expression within pericentric heterochromatin without observing a major effect on pericentric RT (Armstrong et al. 2018). Importantly, almost all RT changes in *H3K9R* mutants within pericentric heterochromatin occur at sites of increased chromatin accessibility and transposon expression suggesting that while a “open”, transcriptionally active environment is

insufficient for replication initiation, it may be necessary (Armstrong et al. 2018). In contrast to *H3K9R* mutation, *Rif1*$^{-}$ mutation results in very few genome-wide changes in chromatin accessibility or transcriptional activity but affects RT at approximately 8% of the genome (Chapter 4,5). Very few of the RT changes observed in *Rif1*$^{-}$ mutants occur within domains of increased chromatin accessibility or transcriptional activity, demonstrating that replication initiation may not require an "open", transcriptionally active environment. To resolve the different conclusions drawn from these data, we propose a model in which multiple independent mechanisms regulate the RT program in metazoans—chromatin accessibility and transcriptional activity regulate RT control independently of chromatin-associated *trans*-acting factors that function at replication origins.

In this model, chromatin accessibility and transcriptional activity contribute to RT control but are insufficient to alter RT programs in the absence of additional molecular changes occurring. For example, in domains of increased chromatin accessibility where RT advances in *H3K9R* mutants, the origin landscape may be altered such that the probability of replication initiation increases; in domains of increased chromatin accessibility where RT is unchanged, the origin landscape may remain unchanged. Importantly, the chromatin landscape is bound by many *trans*-acting factors that modulate RT programs, such as Rif1. An additional, but not mutually exclusive possibility, is that Rif1 localizes normally to pericentric heterochromatin in an *H3K9R* mutant background such that replication initiation remains inhibited by Rif1/PP1 until late S phase despite the increase in chromatin accessibility. Similarly, in *Rif1*$^{-}$ mutants, where domains of altered RT do not display a coincident change in chromatin accessibility or transcriptional activity, redundant molecular mechanisms may promote late RT at the majority of pericentric heterochromatin whereas

domains of altered RT may lack redundancy. To test this model, the next step is to generate genome-wide binding profiles for key contributors to RT control in *H3K9R* and *Rif1*$^{-}$ backgrounds to identify the molecular mechanisms regulating RT control.

Map origins of replication in mutant backgrounds

Current models in the field suggest that replication initiates with a higher probability from euchromatic regions of the genome compared to heterochromatic regions due to a higher density of replication origins in "open" euchromatin relative to "closed" heterochromatin. If an increased density of replication origins is sufficient to advance RT within a particular genomic region, we would expect there to be more licensed origins at regions of the genome that advance RT in *H3K9R* mutants relative to controls. To test this hypothesis, we propose mapping origins of replication in *H3K9R* mutants and controls. Previously, replication origin mapping has been performed on a genome-wide scale using ORC ChIP-seq, short nascent strand sequencing (SNS-seq), replication bubble sequencing (bubble-seq), initiation site sequencing (ini-seq), and Okazaki fragment sequencing (OK-seq) (Besnard et al. 2012; Mesner et al. 2013; Langley et al. 2016; Miotto et al. 2016; Petryk et al. 2016). While these methods provide a genome-wide view of the origin landscape, they were primarily used in cell culture rather than from tissues of a developing organism.

ChIP-seq experiments in *Drosophila* require large tissue input, making them difficult to perform. Cleavage Under Targets and Release Using Nuclease (CUT&RUN) was established by the Henikoff lab as an alternative to ChIP-seq (Skene and Henikoff 2017). In CUT&RUN, ChIP grade antibodies are added directly to your permeabilized sample of interest. Places in the genome where the antibody binds its target protein *in situ* are detected using a MNase-Protein A fusion protein (Skene and Henikoff 2017). Chromatin bound

antibody is cleaved locally by MNase, releasing a DNA fragment into solution for isolation and sequencing (Skene and Henikoff 2017). In a fraction of the time, CUT&RUN provides higher resolution data from lower sequencing depth and lower tissue input than ChIP-seq, making it ideal for studying *Drosophila* tissues (Skene and Henikoff 2017). To initially map origins of replication in *H3K9R* mutants and controls, we propose performing ORC2 CUT&RUN from wing imaginal discs using a ChIP grade ORC2 antibody used previously in ChIP-seq experiments from *Drosophila* tissue (Sher et al. 2012). Comparison of RT profiles generated in *H3K9R* mutants and controls to ORC2 binding profiles will determine whether altered ORC2 binding correlates with altered RT (Armstrong et al. 2018). It is important to note that sites bound by ORC in G1 phase of the cell division cycle are not necessarily the sites from which DNA replication initiates during S phase. This phenomenon is primarily due to the fact that MCM complexes loaded onto chromatin, in and ORC-dependent manner, can slide, often kilobases away from where they were initially loaded (Hyrien 2016). Therefore, to accurately map sites of replication initiation, we propose performing MCM CUT&RUN in *H3K9R* mutants and controls.

If CUT&RUN does not produce high quality data, alternative sequencing approaches can be employed. One such approach, chromatin endogenous cleavage followed by sequencing (ChEC-seq), involves generating an *in vivo* MNase-fusion of your protein of interest to map genome-wide association of that protein (Schmid et al. 2004). The premise of this method is similar to CUT&RUN in that MNase cleaved fragments will be released into solution for isolation and sequencing. However, rather than relying on an antibody to bind your protein of interest, MNase is directly fused to your protein of interest. One major caveat to this method remains that generating fusion proteins can often disrupt the function of your

protein of interest, even though MNase has a low molecular weight (17.6 kDa). This method has been successfully adapted to transcription factors and Rif1, among other factors, suggesting that it is a feasible strategy for profiling genome-wide association of ORC and MCM (Zentner et al. 2015; Hafner et al. 2018).

Identify how Rif1 and H3K9 regulate RT in pericentric heterochromatin

We previously demonstrated that transposon expression and chromatin accessibility at pericentric heterochromatin are perturbed in *H3K9R* mutants (Penke et al. 2016). However, RT of pericentric heterochromatin was only modestly affected in *H3K9R* mutants suggesting that additional mechanisms beyond transcription and chromatin accessibility regulate RT of pericentric heterochromatin (Armstrong et al. 2018). We therefore assessed roles for Rif1 in regulating RT of pericentric heterochromatin and identified that the interaction between Rif1 and PP1 is critical for pericentric heterochromatin late RT (Chapter 4). To further understand this relationship, we propose testing 1) whether Rif1 and H3K9 are dependent on one another for their localization to heterochromatin, 2) whether a genetic interaction exists between Rif1 and H3K9, and 3) whether Rif1 and H3K9 regulate RT of pericentric heterochromatin through distinct mechanisms.

To assess the interdependence of Rif1 and H3K9, we will perform cytological and genomic assays. To determine whether the "hallmarks" of heterochromatin, H3K9me and HP1a, localize properly in *Rif1* mutants, immunofluorescence experiments will be conducted in $Rif1^{-}$, $Rif1^{PP1}$, and wild type control wing imaginal discs for H3K9me2/3 and HP1a. Additionally, we will perform immunofluorescence experiments probing Rif1 in *H3K9R* mutant and *HWT* control wing imaginal discs to identify if H3K9 is required for proper localization of Rif1. To mark pericentric heterochromatin, FISH probes to 359-bp repeat or

to *dodeca* should be included in both experiments. To identify whether loss of Rif1 or H3K9 affects steady state levels of H3K9me/HP1a or Rif1, respectively, western blot analysis will be performed in each respective mutant background for H3K9me2/3, HP1a, and Rif1. Finally, to identify whether Rif1 or H3K9 mutation influences the genome-wide distribution of H3K9me or Rif1, respectively, Rif1 and H3K9me3 CUT&RUN will be performed in each respective mutant background. Together, these experiments will thoroughly identify whether H3K9 and Rif1 are dependent on one another for genome-wide localization to specific chromatin domains.

Previous work from the O'Farrell lab demonstrated that Rif1 promotes the onset of late replication in the early *Drosophila* embryo prior to the onset of mature H3K9me/HP1a-enriched heterochromatic foci (Seller and O'Farrell 2018). However, in the absence of Rif1, H3K9me/HP1a-enrichment still occurs post-mid blastula transition and late replication can still be established, albeit at a later developmental time relative to wild type (Seller and O'Farrell 2018). These data raise the possibility that H3K9 and Rif1 function redundantly to promote late replication of pericentric heterochromatin in *Drosophila*. To identify whether a genetic interaction exists between H3K9 and Rif1, we propose generating *Rif1*$^{-}$*; H3K9R* and *Rif1*$^{-}$; *HWT* double mutants for use in cytological and genomic assays. Viability assays will be performed to identify whether there is synthetic lethality between *Rif1* and *H3K9* mutations. One possibility is that, in the absence of Rif1 and modifiable H3K9, embryonic lethality will result due to defects in heterochromatin establishment. Therefore, immunofluorescence experiments should be performed to assess compaction of DNA and HP1a enrichment during early embryonic development. Furthermore, FAIRE-seq and RNA-seq should be performed in *Rif1*$^{-}$*; H3K9R* and *Rif1*$^{-}$; *HWT* mutants to assess chromatin

accessibility and genome-wide transcription dynamics as a proxy for heterochromatin structure and function. If *Rif1*$^{-}$*; H3K9R* and *Rif1*$^{-}$; *HWT* mutants survive to the third-instar larval stage, RT profiles should be generated to determine if pericentric heterochromatin remains late replicating in the absence of Rif1 and modifiable H3K9. Together these data will identify whether H3K9 and Rif1 function redundantly in regulating heterochromatin structure, function, and RT.

We hypothesize that H3K9 regulates RT primarily through promoting an inaccessible, transcriptionally inert chromatin environment whereas Rif1 regulates RT primarily through its interaction with PP1 and by regulating three-dimensional genome architecture. With the one exception of three-dimensional genome architecture data, we have direct evidence to support all aspects of this hypothesis (Chapter 4) (Penke et al. 2016; Armstrong et al. 2018). To identify whether three-dimensional genome architecture contributes to RT control in *H3K9R* and *Rif1* mutants, we propose assaying the structure of the three-dimensional genome in *H3K9R* and *Rif1* mutants using Hi-C (Lieberman-Aiden et al. 2009).

Identify whether transposon expression is sufficient to induce replication initiation

We previously identified that elevated transposon expression in *H3K9R* mutants is permissive to replication initiation (Armstrong et al. 2018). Despite our experimental attempts to establish causation rather than correlation between replication and gene expression, these data remain correlative. We propose directly testing whether transposon expression is sufficient to advance RT by activating expression of specific transposons and subsequently profiling RT in an otherwise wild type background. Due to the diversity of transposable elements within the *Drosophila* genome, it would be advantageous to

selectively activate different transposons that are part of distinct transposon families. However, the experimental setup here will be described regarding only one transposon family.

Previous work from our lab identified that *gypsy* transposons, located in the pericentric heterochromatin of Chromosome X, are upregulated and mobilize within the genome of *H3K9R* mutants (Penke et al. 2016). To directly test whether *gypsy* expression is sufficient to advance RT in an otherwise wild type background, we will employ a genetic strategy to both activate and suppress *gypsy* expression within the follicle cells of the adult ovary. The *flamenco* locus within the pericentric heterochromatin of Chromosome X has previously been shown to regulate maternal expression of *gypsy* in follicle cells (Prud'homme et al. 1995). A permissive genotype (homozygous for the *flamP* allele of *flamenco*) allows for robust *gypsy* expression within the follicle cells whereas a dominant restrictive genotype (at least one *flamR* allele of *flamenco*) inhibits *gypsy* expression within the follicle cells of the female ovary (Touret et al. 2014). We propose profiling RT in *flamP*/*flamP*, *flamR*/*flamR*, *flamP*/+, *flamR*/+, and wild type ovarian follicle cells to identify if altered *gypsy* expression is sufficient to advance RT of Chromosome X pericentric heterochromatin. Furthermore, to identify how each *flamenco* genotype influences genome-wide transcription dynamics, we will also perform RNA-seq in these genetic backgrounds.

While the genetic strategy for altering *gypsy* expression is specific to the *gypsy* transposon family, additional strategies can be employed that are more broadly applicable, such as site-specific targeting of transcriptional activators. To easily screen many different transposons concurrently, we propose a UAS/Gal4 based approach, where UAS-driven transgenes containing a transposon of interest are incorporated at a specific genomic location.

We recommend conducting this experiment at multiple genomic sites within distinct chromatin environments (i.e. euchromatic or heterochromatic). Upon activation of the transgenic transposon, pools of S phase cells will be isolated from wing imaginal discs using FACS, and a series of experimental and control qPCR reactions will be performed to identify the relative RT of the transgenic locus. One caveat to this method, however, is that the transcriptional environment used for expression of transgenic transposons is not identical to that of the endogenous machinery. However, if *gypsy* expression using the genetic tools of the *flamenco* locus is sufficient to alter RT, transgenic *gypsy* expression can be induced and directly compared to the genetic *gypsy* expression system as a control experiment to assess the quality and potential utility of a transgenic transposon experimental system.

Investigate the relationship between RT and genome instability in mutant backgrounds

There is an established relationship between RT and genomic instability—late replicating regions of the genome tend to experience an increased mutation rate relative to early replicating regions of the genome (Watanabe and Maekawa 2010). However, how individual histone residues that structurally contribute to euchromatic and heterochromatic domains function to maintain genome integrity during DNA replication is largely unknown. Previous work from the MacAlpine lab demonstrated that while the H4K20 mono-methyltransferase PR-Set7 was dispensable for proper RT control, it was required to maintain genomic integrity specifically at late replicating regions of the genome demonstrating that RT and DNA damage are mechanistically separable (Li et al. 2016). We propose profiling DNA damage on a genome-wide scale in histone mutants to interrogate roles for histone residues in maintaining genome stability. Furthermore, direct comparison of

RT data with genome instability data may provide a greater understanding for how altered RT programs contribute to genomic instability in the context of a developing organism.

To profile DNA damage on a genome-wide scale, we propose performing CUT&RUN using a monoclonal antibody developed in the Sekelsky lab that recognizes a phosphorylated form of the histone variant H2Av (γH2Av) (Lake et al. 2013). In *Drosophila*, a DNA double-stranded break is initially marked by a phosphorylated form of the histone variant H2Av, and recognition of γH2Av as a proxy for DNA double-stranded breaks has become the gold standard for DNA damage detection. Preliminary cytological studies from our lab show elevated genomic instability in *H3K9R*, *H4K16R*, and *H3K56R* mutant imaginal disc tissue (Appendix 1). We therefore propose initially performing γH2Av CUT&RUN from wing imaginal discs in *H3K9R*, *H4K16R*, and *H3K56R* mutants and their respective controls. Additionally, because the H4K20 mono-methyltransferase PR-Set7 has been implicated in maintaining genomic integrity of late replicating sequences, we also propose performing γH2Av CUT&RUN from wing imaginal discs of *H4K20R* mutants. As genome-wide RT datasets have only been generated for *H3K9R* and *H4K16R* mutants, genome-wide RT datasets for *H3K56R* and *H4K20R* mutants should be generated to allow for direct comparison of RT and genome instability.

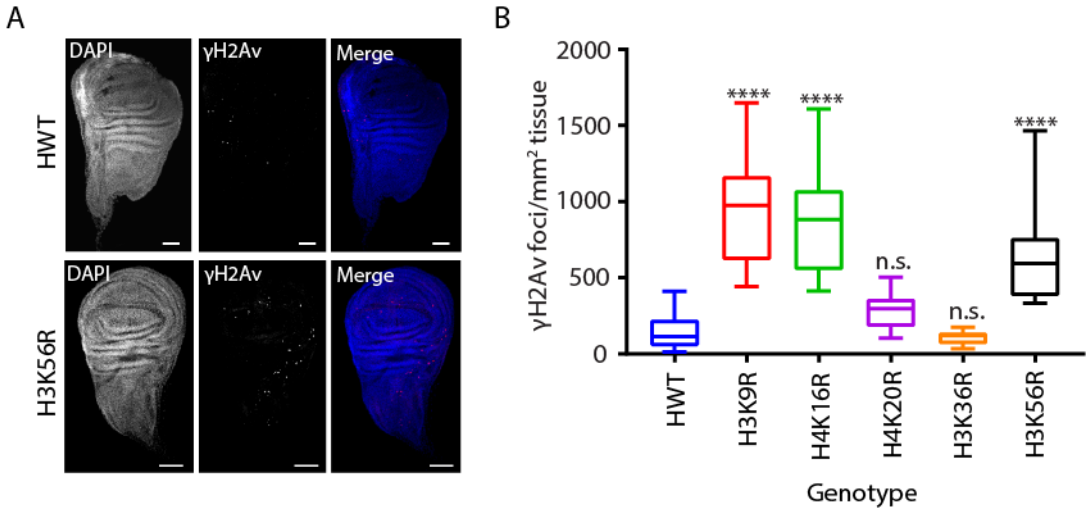
APPENDIX 1

DNA repair occurs in a dynamic chromatin environment that provides a barrier to both the detection and downstream repair of DNA lesions. Evidence supports a model where chromatin surrounding a DNA lesion must detect DNA damage, allow remodeling for repair, and restore its original chromatin structure (Chiolo et al. 2011; Polo and Jackson 2011; Price and D'Andrea 2013; Ayrapetov et al. 2014; Burgess et al. 2014). Constitutive heterochromatin, characterized by H3K9me2/3, provides a challenging environment for DNA repair (Jenuwein and Allis 2001). Studies in *Drosophila* and mammalian cell culture demonstrate that the activity of the H3K9 methyltransferase Suv39h1 (Su(var)3-9 in *Drosophila*) promotes activation of the checkpoint kinase ATM following DNA damage and activates relocation of breaks outside of heterochromatin for repair (Chiolo et al. 2011; Ayrapetov et al. 2014). Suv39h1 mutants show defects in heterochromatin stability with elevated DNA damage, mitotic checkpoint activation, and repeat instability (Peng and Karpen 2006; Peng and Karpen 2009; Ayrapetov et al. 2014). Additional histone residues are implicated in non-homologous end-joining repair of DNA DSBs including a proposed requirement for dynamic H4K20me2 and H4K16ac for recruitment of 53BP1 (Li et al. 2010; Hsiao and Mizzen 2013; Dulev et al. 2014; Tuzon et al. 2014). H4K16ac has been hypothesized to function upstream of 53BP1 recruitment through direct interaction with the histone 2A variant, H2A.X (Li et al. 2010). A H4K20me1 to H4K20me2 transition is hypothesized to function in regulating 53BP1 nucleation, although the requirement for these methylation states, rather than the writers, in this process remains unclear (Tuzon et al. 2014). Additionally, H3K36 is implicated in homologous recombination repair of a DSB. Reduction of H3K36me3 levels results in reduced homologous recombination DSB repair

efficiency and an increased frequency of deletion repair products (Pfister et al. 2014). Cancers deficient for the H3K36 methyltransferase SETD2 exhibit a wide range of mutations, further suggesting a role for H3K36 in maintenance of genome stability (Sato et al. 2013; Zhu et al. 2014). Finally, H3K56 acetylation is implicated in the repair of DNA legions generated during DNA replication as yeast cells devoid of H3K56ac, either through *H3K56R* mutation or mutation of the yeast acetyltransferase *rtt109*, are sensitive to genotoxic agents associated with replication stress (Masumoto et al. 2005; Driscoll et al. 2007; Wurtele et al. 2012).

A feature of genome instability is the generation of DNA double-stranded breaks (DSBs). In *Drosophila*, DSBs recruit a phosphorylated form of H2Av (histone 2A variant). To identify potential roles for histone residues in genome stability, we performed γH2Av staining in wing discs of a panel of histone mutants. Damage was scored as γH2Av foci normalized to area. While H3K9, H4K16, and H3K56R mutants showed a dramatic increase in γH2Av signal compared to HWT, H4K20 mutants only showed a moderate increase and H3K36 mutants showed no increase in γH2Av signal, respectively, indicating genomic instability in H3K9, H4K16, and H3K56R mutants (Figure A1.A,B).

Figure A1. H3K9, H4K16, and H3K56 promote genome stability. A) Representative *HWT* (top) and *H3K56R* (bottom) wing imaginal discs stained for γH2Av (red) and DAPI (blue). **B)** The average number of γH2Av foci per mm^2 of wing imaginal disc tissue for all tissues analyzed for *HWT* (blue), *H3K9R* (red), *H4K16R* (green), *H4K20R* (purple), *H3K36R* (orange), and *H3K56R* (black). Images were thresholded using the "intermodes" automatic threshold setting in the ImageJ software. One-way ANOVA was performed. Significance is indicated for comparisons to *HWT* control; **** = $p < 0.0001$.



APPENDIX 2

Histone PTMs are associated with origins across species and have been strongly correlated with the time at which a particular DNA sequence replicates during S phase. Euchromatin tends to replicate early during S phase whereas heterochromatin tends to replicate late during S phase (Schwaiger et al. 2009; Unnikrishnan et al. 2010; Eaton et al. 2011; Petruk et al. 2013; Lubelsky et al. 2014). Current models suggest that replication initiation factors are recruited to euchromatin with a higher probability than heterochromatin, most likely contributing to their differential RT (Mantiero et al. 2011; Collart et al. 2013; Das et al. 2015). Previous efforts have sought to identify functions for histone residues in the regulation of RT through modulation of writer, reader, or eraser proteins, with select studies summarized here.

Although H3K9me is enriched at late replicating sequences, mutation of H3K9 writers and readers provides contradictory conclusions regarding the role of H3K9 in modulating RT. In *S. pombe*, mutation of Swi6, the HP1 homologue and H3K9me reader results in a severe delay in RT; however, mutation of Clr4, the H3K9 methyltransferase, results in advanced RT (Li et al. 2013). Furthermore, knockdown of HP1 in *Drosophila* cells shows a biphasic effect on RT where pericentromeric heterochromatin RT advances and HP1-bound sites on chromosome arms have delayed RT (Schwaiger et al. 2010). Roles for H4K20 in RT are also unclear. Mammalian studies suggest that localization of the H4K20me1 writer, PR-Set7, to replication foci and its subsequent degradation is required to prevent aberrant S phase phenotypes (Houston et al. 2008; Tardat et al. 2010). However, studies in *Drosophila* cells demonstrate that PR-Set7 is not required for origin activation or proper RT programs throughout the genome suggesting that H4K20me states do not function

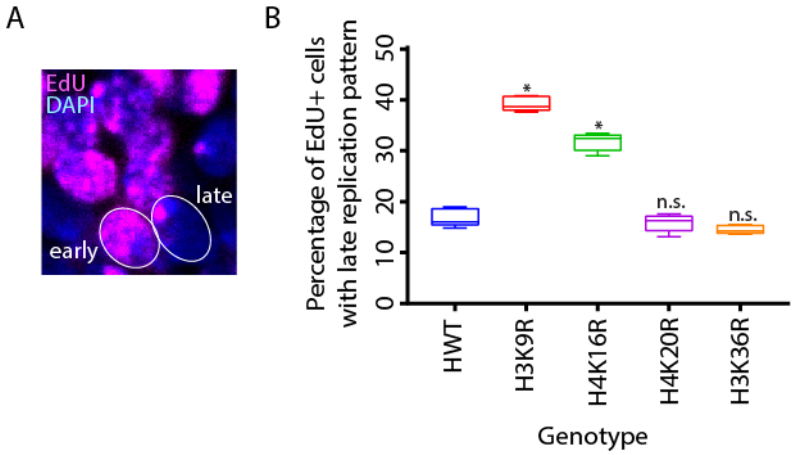
in RT (Li et al. 2016). Moreover, H4K20A studies in our lab showed no requirement for H4K20 modification in DNA replication in the ovary, supporting a lack of functional requirement for H4K20 in DNA replication (McKay et al. 2015).Together, these contrasting results preclude interpretation of H3K9 and H4K20's roles in RT.

Although informative, histone PTM correlation studies cannot determine direct roles for histone residues in RT control. Studies in Drosophila show enrichment of H4ac marks, including H4K16ac, at sites of replication initiation in ovarian follicle cells (Aggarwal and Calvi 2004). Forced de-acetylation at these sites is sufficient to impair origin activation efficiency supporting a role for H4ac in origin activation (Aggarwal and Calvi 2004; Liu et al. 2012). Furthermore, dosage compensation in *Drosophila* males is mediated through hyper-H4K16ac of the male X chromosome (a chromatin state which promotes early replication and high transcriptional activity); mutation of H4K16ac writers MSL2 and MOF is sufficient to reduce origin activation on the male X chromosome, resulting in a loss of early RT (Lubelsky et al. 2014). However, in each of these contexts, whether H4K16ac is directly required for early origin activation remains unknown.

In *Drosophila*, pericentromeric heterochromatin clusters into a nuclear structure termed the chromocenter. The chromocenter can be visualized as a DAPI-bright focus while euchromatin stains dimly with DAPI. EdU pulse labelling patterns define cells in early and late S phase, with early S phase cells identified by euchromatic EdU labeling and late S phase visualized by heterochromatic EdU labeling (Figure A2A). Quantifying the relative proportions of EdU-positive cells with either early or late replication patterns can serve as a proxy for RT changes. To identify histone mutations that disrupt RT, we quantified early and late S phase patterns in cells that were pulse labeled for 60-minutes with EdU. We found that

H3K9R and *H4K16R* mutants display an increased frequency of late replicating cells while *H4K20R* and *H3K36R* mutants are not different than the control (Figure A2B). From these data, I conclude that H4K16 and H3K9, but not H4K20 or H3K36R, are required for proper regulation of the RT program.

Figure A2. Replication pattern analysis in *Drosophila* histone mutants. A) Eye imaginal discs pulse labeled for 1hr with EdU (magenta) and stained for DNA (blue; DAPI). Representative early and late patterns are indicated. D) Percentage of EdU+ cells with early or late EdU incorporation patterns for *HWT* (blue), *H3K9R* (red), *H4K16R* (green), *H4K20R* (purple), and *H3K36R* (orange) genotypes. Chi squared test was performed (* = $p < 0.01$).



REFERENCES

Abyzov A, Urban AE, Snyder M, Gerstein M. 2011. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome research* **21**: 974-984.

Aggarwal BD, Calvi BR. 2004. Chromatin regulates origin activity in *Drosophila* follicle cells. *Nature* **430**: 372-376.

Almeida R, Fernández-Justel JM, Santa-María C, Cadoret J-C, Cano-Aroca L, Lombraña R, Herranz G, Agresti A, Gómez M. 2018. Chromatin conformation regulates the coordination between DNA replication and transcription. *Nature Communications* **9**: 1590.

Alver RC, Chadha GS, Gillespie PJ, Blow JJ. 2017. Reversal of DDK-Mediated MCM Phosphorylation by Rif1-PP1 Regulates Replication Initiation and Replisome Stability Independently of ATR/Chk1. *Cell reports* **18**: 2508-2520.

Amiel A, Litmanovitch T, Lishner M, Mor A, Gaber E, Tangi I, Fejgin M, Avivi L. 1998. Temporal differences in replication timing of homologous loci in malignant cells derived from CML and lymphoma patients. *Genes Chromosomes Cancer* **22**: 225-231.

Andreyeva EN, Bernardo TJ, Kolesnikova TD, Lu X, Yarinich LA, Bartholdy BA, Guo X, Posukh OV, Healton S, Willcockson MA et al. 2017. Regulatory functions and chromatin loading dynamics of linker histone H1 during endoreplication in Drosophila. *Genes Dev* **31**: 603-616.

Andreyeva EN, Kolesnikova TD, Belyaeva ES, Glaser RL, Zhimulev IF. 2008. Local DNA underreplication correlates with accumulation of phosphorylated H2Av in the Drosophila melanogaster polytene chromosomes. *Chromosome Research* **16**: 851-862.

Armstrong RL, Duronio RJ. Cell cycle analysis in H4K16R diploid wing imaginal disc cells. *(unpublished)*.

Armstrong RL, Penke TJR, Strahl BD, Matera AG, McKay DJ, MacAlpine DM, Duronio RJ. 2018. Chromatin conformation and transcriptional activity are permissive regulators of DNA replication initiation in Drosophila. *Genome research* doi:10.1101/gr.239913.118.

Ayrapetov MK, Gursoy-Yuzugullu O, Xu C, Xu Y, Price BD. 2014. DNA double-strand breaks promote methylation of histone H3 on lysine 9 and transient formation of repressive chromatin. *PNAS* **111**: 9169-9174.

Becker JS, McCarthy RL, Sidoli S, Donahue G, Kaeding KE, He Z, Lin S, Garcia BA, Zaret KS. 2017. Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. *Mol Cell* **68**: 1023-1037.e1015.

Bell O, Schwaiger M, Oakeley EJ, Lienert F, Beisel C, Stadler MB, Schubeler D. 2010. Accessibility of the Drosophila genome discriminates PcG repression, H4K16 acetylation and replication timing. *Nature structural & molecular biology* **17**: 894-900.

Bell SP. 2017. Rethinking origin licensing. *eLife* **6**: e24052.

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

Belyaeva ES, Zhimulev IF, Volkova EI, Alekseyenko AA, Moshkin YM, Koryakov DE. 1998. Su(UR)ES: a gene suppressing DNA underreplication in intercalary and pericentric heterochromatin of Drosophila melanogaster polytene chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 7532-7537.

Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, Marin J-M, Lemaitre J-M. 2012. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature structural & molecular biology* **19**: 837-844.

Black Joshua C, Manning Amity L, Van Rechem C, Kim J, Ladd B, Cho J, Pineda Cristiana M, Murphy N, Daniels Danette L, Montagna C et al. 2013. KDM4A Lysine Demethylase Induces Site-Specific Copy Gain and Rereplication of Regions Amplified in Tumors. *Cell* **154**: 541-555.

Bolstad BM. 2016. preprocessCore: A collection of pre-processing functions. *R package version 1360*.

Breger KS, Smith L, Thayer MJ. 2005. Engineering translocations with delayed replication: evidence for cis control of chromosome replication timing. *Human molecular genetics* **14**: 2813-2827.

Breger KS, Smith L, Turker MS, Thayer MJ. 2004. Ionizing radiation induces frequent translocations with delayed replication and condensation. *Cancer Res* **64**: 8231-8238.

Buonomo SBC, Wu Y, Ferguson D, de Lange T. 2009. Mammalian Rif1 contributes to replication stress survival and homology-directed repair. *The Journal of Cell Biology* **187**: 385-398.

Burgess RC, Burman B, Kruhlak MJ, Misteli T. 2014. Activation of DNA Damage Response Signaling by Condensed Chromatin. *Cell reports* **9**: 1-15.

Buscaino A, Kocher T, Kind JH, Holz H, Taipale M, Wagner K, Wilm M, Akhtar A. 2003. MOF-regulated acetylation of MSL-3 in the Drosophila dosage compensation complex. *Mol Cell* **11**: 1265-1277.

Calvi BR, Lilly MA, Spradling AC. 1998. Cell cycle control of chorion gene amplification. *Genes & development* **12**: 734-744.

Canzio D, Chang EY, Shankar S, Kuchenbecker KM, Simon MD, Madhani HD, Narlikar GJ, Al-Sady B. 2011. Chromodomain-Mediated Oligomerization of HP1 Suggests a Nucleosome-Bridging Mechanism for Heterochromatin Assembly. *Molecular Cell* **41**: 67-81.

Cayrou C, Coulombe P, Vigneron A, Stanojcic S, Ganier O, Peiffer I, Rivals E, Puy A, Laurent-Chabalier S, Desprat R et al. 2011. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome research* **21**: 1438-1449.

Celniker SE, Dillon LAL, Gerstein MB, Gunsalus KC, Henikoff S, Karpen GH, Kellis M, Lai EC, Lieb JD, MacAlpine DM et al. 2009. Unlocking the secrets of the genome. *Nature* **459**: 927.

Chiolo I, Minoda A, Colmenares SU, Polyzos A, Costes SV, Karpen GH. 2011. Double-Strand Breaks in Heterochromatin Move Outside of a Dynamic HP1a Domain to Complete Recombinational Repair. *Cell* **144**: 732-744.

Coleman TR, Carpenter PB, Dunphy WG. 1996. The Xenopus Cdc6 protein is essential for the initiation of a single round of DNA replication in cell-free extracts. *Cell* **87**: 53-63.

Collart C, Allen GE, Bradshaw CR, Smith JC, Zegerman P. 2013. Titration of Four Replication Factors Is Essential for the Xenopus laevis Midblastula Transition. *Science* **341**: 893-896.

Comoglio F, Paro R. 2014. Combinatorial Modeling of Chromatin Features Quantitatively Predicts DNA Replication Timing in Drosophila. *PLOS Computational Biology* **10**: e1003419.

Copur O, Gorchakov A, Finkl K, Kuroda MI, Muller J. 2018. Sex-specific phenotypes of histone H4 point mutants establish dosage compensation as the critical function of H4K16 acetylation in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **115**: 13336-13341.

Cornacchia D, Dileep V, Quivy JP, Foti R, Tili F, Santarella-Mellwig R, Antony C, Almouzni G, Gilbert DM, Buonomo SBC. 2012. Mouse Rif1 is a key regulator of the replication-timing programme in mammalian cells. *The EMBO Journal* **31**: 3678-3690.

Cui P, Ding F, Lin Q, Zhang L, Li A, Zhang Z, Hu S, Yu J. 2012. Distinct contributions of replication and transcription to mutation rate variation of human genomes. *Genomics Proteomics Bioinformatics* **10**: 4-10.

Cytron S, Stepnov E, Bounkin I, Mashevich M, Dotan A, Avivi L. 2011. Epigenetic analyses in blood cells of men suspected of prostate cancer predict the outcome of biopsy better than serum PSA levels. *Clin Epigenetics* **2**: 383-388.

Das SP, Borrman T, Liu VWT, Yang SC-H, Bechhoefer J, Rhind N. 2015. Replication timing is regulated by the number of MCMs loaded at origins. *Genome research* **25**: 1886-1892.

Davé A, Cooley C, Garg M, Bianchi A. 2014. Protein Phosphatase 1 Recruitment by Rif1 Regulates DNA Replication Origin Firing by Counteracting DDK Activity. *Cell reports* **7**: 53-61.

De Lucia F, Ni J, Vaillant C, Sun F. 2005. HP1 modulates the transcription of cell-cycle regulators in Drosophila melanogaster. *Nucleic acids research* **33**: 2852-2858.

De S, Michor F. 2011. DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. *Nat Biotech* **29**: 1103-1108.

Dean P, Jett J. 1974. Mathematical analysis of DNA distributions from flow microfluorometry. *Journal of Cell Biology* **60**: 523-527.

Delgado S, Gómez M, Bird A, Antequera F. 1998. Initiation of DNA replication at CpG islands in mammalian chromosomes. *The EMBO Journal* **17**: 2426-2435.

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

Dileep V, Ay F, Sima J, Vera DL, Noble WS, Gilbert DM. 2015. Topologically associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication-timing program. *Genome research* **25**: 1104-1113.

Dileep V, Gilbert DM. 2018. Single-cell replication profiling to measure stochastic variation in mammalian replication timing. *Nature Communications* **9**: 427.

Dimitrova DS, Gilbert DM. 1999. The spatial position and replication timing of chromosomal domains are both established in early G1 phase. *Molecular cell* **4**: 983-993.

Donley N, Thayer MJ. 2013. DNA replication timing, genome stability and cancer: Late and/or delayed DNA replication timing is associated with increased genomic instability. *Seminars in Cancer Biology* **23**: 80-89.

Driscoll R, Hudson A, Jackson SP. 2007. Yeast Rtt109 Promotes Genome Stability by Acetylating Histone H3 on Lysine 56. *Science* **315**: 649-652.

Du Q, Bert SA, Armstrong NJ, Caldon CE, Song JZ, Nair SS, Gould CM, Luu P-L, Peters T, Khoury A et al. 2019. Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. *Nature Communications* **10**: 416.

Dulev S, Tkach J, Lin S, Batada NN. 2014. SET8 methyltransferase activity during the DNA double-strand break response is required for recruitment of 53BP1. *EMBO reports* doi:10.15252/embr.201439434.

Duzdevich D, Warner Megan D, Ticau S, Ivica Nikola A, Bell Stephen P, Greene Eric C. 2015. The Dynamics of Eukaryotic Replication Initiation: Origin Specificity, Licensing, and Firing at the Single-Molecule Level. *Molecular Cell* **58**: 483-494.

Eaton ML, Prinz JA, MacAlpine HK, Tretyakov G, Kharchenko PV, MacAlpine DM. 2011. Chromatin signatures of the *Drosophila* replication program. *Genome research* **21**: 164-174.

Edgar BA, Orr-Weaver TL. 2001. Endoreplication cell cycles: more for less. *Cell* **105**: 297-306.

Edgar BA, Zielke N, Gutierrez C. 2014. Endocycles: a recurrent evolutionary innovation for post-mitotic cell growth. *Nature reviews Molecular cell biology* **15**: 197-210.

Evrin C, Clarke P, Zech J, Lurz R, Sun J, Uhle S, Li H, Stillman B, Speck C. 2009. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proceedings of the National Academy of Sciences* **106**: 20240-20245.

Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ et al. 2010. Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell* **143**: 212-224.

Foti R, Gnan S, Cornacchia D, Dileep V, Bulut-Karslioglu A, Diehl S, Buness A, Klein Felix A, Huber W, Johnstone E et al. 2016. Nuclear Architecture Organized by Rif1 Underpins the Replication-Timing Program. *Molecular Cell* **61**: 1-14.

Fox DT, Duronio RJ. 2013. Endoreplication and polyploidy: insights into development and disease. *Development* **140**: 3-12.

Fox M. 1980. A Model for the Computer Analysis of Synchronous DNA Distribution. *Cytometry* **1**: 71-77.

Fritz A, Sinha S, Marella N, Berezney R. 2013. Alterations in replication timing of cancer-related genes in malignant human breast cancer cells. *Journal of Cellular Biochemistry* **114**: 1074-1083.

Gelbart ME, Larschan E, Peng S, Park PJ, Kuroda MI. 2009. Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nature Structural &Amp; Molecular Biology* **16**: 825.

Ghavi-Helm Y, Jankowski A, Meiers S, Viales RR, Korbel JO, Furlong EEM. 2019. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nature genetics* **51**: 1272-1282.

Gindin Y, Valenzuela MS, Aladjem MI, Meltzer PS, Bilke S. 2014. A chromatin structure-based model accurately predicts DNA replication timing in human cells. *Mol Syst Biol* **10**: 722-722.

Glozak MA, Sengupta N, Zhang X, Seto E. 2005. Acetylation and deacetylation of non-histone proteins. *Gene* **363**: 15-23.

Goldman M, Holmquist G, Gray M, Caston L, Nag A. 1984. Replication timing of genes and middle repetitive sequences. *Science* **224**: 686-692.

Grant GD, Kedziora KM, Limas JC, Cook JG, Purvis JE. 2018. Accurate delineation of cell cycle phase transitions in living cells with PIP-FUCCI. *Cell cycle (Georgetown, Tex)* **17**: 2496-2516.

Gros J, Kumar C, Lynch G, Yadav T, Whitehouse I, Remus D. 2015. Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. *Molecular cell* **60**: 797-807.

Günesdogan U, Jäckle H, Herzig A. 2010. A genetic system to asses *in vivo* the functions of histones and histone modifications in higher eukaryotes. *EMBO reports* **11**: 772-776.

Hafner L, Lezaja A, Zhang X, Lemmens L, Shyian M, Albert B, Follonier C, Nunes JM, Lopes M, Shore D et al. 2018. Rif1 Binding and Control of Chromosome-Internal DNA Replication Origins Is Limited by Telomere Sequestration. *Cell reports* **23**: 983-992.

Hammond MP, Laird CD. 1985. Control of DNA replication and spatial distribution of defined DNA sequences in salivary gland cells of Drosophila melanogaster. *Chromosoma* **91**: 279-286.

Hamperl S, Cimprich KA. 2016. Conflict Resolution in the Genome: How Transcription and Replication Make It Work. *Cell* **167**: 1455-1467.

Hannibal RL, Chuong EB, Rivera-Mulia JC, Gilbert DM, Valouev A, Baker JC. 2014. Copy number variation is a fundamental aspect of the placental genome. *PLoS genetics* **10**: e1004290.

Hayano M, Kanoh Y, Matsumoto S, Renard-Guillet C, Shirahige K, Masai H. 2012. Rif1 is a global regulator of timing of replication origin firing in fission yeast. *Genes & Development* **26**: 137-150.

Haynes KA, Gracheva E, Elgin SCR. 2007. A Distinct Type of Heterochromatin Within Drosophila melanogaster Chromosome 4. *Genetics* **175**: 1539-1542.

Heinz KS, Casas-Delucchi CS, Török T, Cmarko D, Rapp A, Raska I, Cardoso M C. 2018. Peripheral re-localization of constitutive heterochromatin advances its replication timing and impairs maintenance of silencing marks. *Nucleic acids research* doi:10.1093/nar/gky368: gky368-gky368.

Hilfiker A, Hilfiker-Kleiner D, Pannuti A, Lucchesi JC. 1997. mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in Drosophila. *The EMBO Journal* **16**: 2054-2060.

Hiraga S-i, Alvino GM, Chang F, Lian H-y, Sridhar A, Kubota T, Brewer BJ, Weinreich M, Raghuraman MK, Donaldson AD. 2014. Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse Cdc7-mediated phosphorylation of the MCM complex. *Genes & Development* **28**: 372-383.

Hiraga Si, Ly T, Garzón J, Hořejší Z, Ohkubo Yn, Endo A, Obuse C, Boulton SJ, Lamond AI, Donaldson AD. 2017. Human RIF1 and protein phosphatase 1 stimulate DNA replication origin licensing but suppress origin activation. *EMBO reports* doi:10.15252/embr.201641983.

Hiratani I, Ryba T, Itoh M, Rathjen J, Kulik M, Papp B, Fussner E, Bazett-Jones DP, Plath K, Dalton S et al. 2010. Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. *Genome research* **20**: 155-169.

Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, Chang C-W, Lyou Y, Townes TM, Schübeler D, Gilbert DM. 2008. Global Reorganization of Replication Domains During Embryonic Stem Cell Differentiation. *PLOS Biology* **6**: e245.

Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer BD, George RA, Svirskas R et al. 2015. The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome research* **25**: 445-458.

Houston SI, McManus KJ, Adams MM, Sims JK, Carpenter PB, Hendzel MJ, Rice JC. 2008. Catalytic Function of the PR-Set7 Histone H4 Lysine 20 Monomethyltransferase Is Essential for Mitotic Entry and Genomic Stability. *Journal of Biological Chemistry* **283**: 19478-19488.

Hsiao K-Y, Mizzen CA. 2013. Histone H4 deacetylation facilitates 53BP1 DNA damage signaling and double-strand break repair. *Journal of Molecular Cell Biology* **5**: 157-165.

Hua BL, Bell GW, Kashevsky H, Von Stetina JR, Orr-Weaver TL. 2018. Dynamic changes in ORC localization and replication fork progression during tissue differentiation. *BMC genomics* **19**: 623.

Hua BL, Orr-Weaver TL. 2017. DNA Replication Control During Drosophila Development: Insights into the Onset of S Phase, Replication Initiation, and Fork Progression. *Genetics* **207**: 29-47.

Huang J, Berger SL. 2008. The emerging field of dynamic lysine methylation of non-histone proteins. *Current Opinion in Genetics & Development* **18**: 152-158.

Hyrien O. 2016. How MCM loading and spreading specify eukaryotic DNA replication initiation sites. *F1000Res* **5**: F1000 Faculty Rev-2063.

Jenuwein T, Allis CD. 2001. Translating the Histone Code. *Science* **293**: 1074-1080.

Joyce EF, Williams BR, Xie T, Wu Ct. 2012. Identification of Genes That Promote or Antagonize Somatic Homolog Pairing Using a High-Throughput FISH–Based Screen. *PLoS genetics* **8**: e1002667.

Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. 2004. The UCSC Table Browser data retrieval tool. *Nucleic acids research* **32**: D493-D496.

Kharchenko PV, Aleksyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sobo PJ, Larschan E, Gorchakov AA, Gu T et al. 2011. Comprehensive analysis of the chromatin landscape in Drosophila. *Nature* **471**: 480-485.

Kolesnikova TD, Goncharov FP, Zhimulev IF. 2018. Similarity in replication timing between polytene and diploid cells is associated with the organization of the Drosophila genome. *PloS one* **13**: e0195207.

Kolesnikova TD, Posukh OV, Andreyeva EN, Bebyakina DS, Ivankin AV, Zhimulev IF. 2013. Drosophila SUUR protein associates with PCNA and binds chromatin in a cell cycle-dependent manner. *Chromosoma* **122**: 55-66.

Kondo S, Ueda R. 2013. Highly Improved Gene Targeting by Germline-Specific Cas9 Expression in Drosophila. *Genetics* **195**: 715-721.

Koren A, Handsaker RE, Kamitaki N, Karlić R, Ghosh S, Polak P, Eggan K, McCarroll SA. 2014. Genetic variation in human DNA replication timing. *Cell* **159**: 1015-1026.

Koren A, Polak P, Nemesh J, Michaelson Jacob J, Sebat J, Sunyaev Shamil R, McCarroll Steven A. 2012. Differential Relationship of DNA Replication Timing to Different Forms of Human Mutation and Variation. *American Journal of Human Genetics* **91**: 1033-1040.

Korenstein-Ilan A, Amiel A, Lalezari S, Lishner M, Avivi L. 2002. Allele-specific replication associated with aneuploidy in blood cells of patients with hematologic malignancies. *Cancer Genetics and Cytogenetics* **139**: 97-103.

Kuroda MI, Hilfiker A, Lucchesi JC. 2016. Dosage Compensation in Drosophila—a Model for the Coordinate Regulation of Transcription. *Genetics* **204**: 435-450.

Lake CM, Holsclaw JK, Bellendir SP, Sekelsky J, Hawley RS. 2013. The Development of a Monoclonal Antibody Recognizing the Drosophila melanogaster Phosphorylated Histone H2A Variant (γ-H2AV). *G3: Genes|Genomes|Genetics* **3**: 1539-1543.

Lander ES Linton LM Birren B Nusbaum C Zody MC Baldwin J Devon K Dewar K Doyle M FitzHugh W et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.

Langley AR, Gräf S, Smith JC, Krude T. 2016. Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq). *Nucleic acids research* **44**: 10230-10247.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**: R25.

Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, Agard DA, Redding S, Narlikar GJ. 2017. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**: 236-240.

Lee HO, Davidson JM, Duronio RJ. 2009. Endoreplication: polyploidy with purpose. *Genes & Development* **23**: 2461-2477.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Li P-C, Green MD, Forsburg SL. 2013. Mutations Disrupting Histone Methylation Have Different Effects on Replication Timing in S. pombe Centromere. *PloS one* **8**: e61464.

Li X, Corsa CAS, Pan PW, Wu L, Ferguson D, Yu X, Min J, Dou Y. 2010. MOF and H4 K16 Acetylation Play Important Roles in DNA Damage Repair by Modulating Recruitment of DNA Damage Repair Protein Mdc1. *Molecular and Cellular Biology* **30**: 5335-5347.

Li Y, Armstrong RL, Duronio RJ, MacAlpine DM. 2016. Methylation of histone H4 lysine 20 by PR-Set7 ensures the integrity of late replicating sequence domains in Drosophila. *Nucleic acids research* doi:10.1093/nar/gkw333.

Liang C, Weinreich M, Stillman B. 1995. ORC and Cdc6p interact and determine the frequency of initiation of DNA replication in the genome. *Cell* **81**: 667-676.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO et al. 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**: 289-293.

Lilly MA, Duronio RJ. 2005. New insights into cell cycle control from the Drosophila endocycle. *Oncogene* **24**: 2765-2775.

Lilly MA, Spradling AC. 1996. The Drosophila endocycle is controlled by Cyclin E and lacks a checkpoint ensuring S-phase completion. *Genes Dev* **10**: 2514-2526.

Litmanovitch T, Altaras MM, Dotan A, Avivi L. 1998. Asynchronous replication of homologous alpha-satellite DNA loci in man is associated with nondisjunction. *Cytogenet Cell Genet* **81**: 26-35.

Liu J, McConnell K, Dixon M, Calvi BR. 2012. Analysis of model replication origins in *Drosophila* reveals new aspects of the chromatin landscape and its relationship to origin activity and the prereplicative complex. *Molecular biology of the cell* **23**: 200-212.

Lubelsky Y, Prinz JA, DeNapoli L, Li Y, Belsky JA, MacAlpine DM. 2014. DNA replication and transcription programs respond to the same chromatin cues. *Genome research* **24**: 1102-1114.

Lucchesi JC. 1998. Dosage compensation in flies and worms: the ups and downs of X-chromosome regulation. *Current Opinion in Genetics & Development* **8**: 179-184.

Ma J, Weake VM. 2014. Affinity-based Isolation of Tagged Nuclei from Drosophila Tissues for Gene Expression Analysis. doi:doi:10.3791/51418: e51418.

MacAlpine DM, Rodriguez HK, Bell SP. 2004. Coordination of replication and transcription along a *Drosophila* chromosome. *Genes & Development* **18**: 3094-3105.

MacAlpine HK, Gordân R, Powell SK, Hartemink AJ, MacAlpine DM. 2010. Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome research* **20**: 201-211.

Makunin IV, Volkova EI, Belyaeva ES, Nabirochkina EN, Pirrotta V, Zhimulev IF. 2002. The Drosophila suppressor of underreplication protein binds to late-replicating regions of polytene chromosomes. *Genetics* **160**: 1023-1034.

Mantiero D, Mackenzie A, Donaldson A, Zegerman P. 2011. Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *The EMBO Journal* **30**: 4805-4814.

Maqbool SB, Mehrotra S, Kolpakas A, Durden C, Zhang B, Zhong H, Calvi BR. 2010. Dampened activity of E2F1-DP and Myb-MuvB transcription factors in Drosophila endocycling cells. *J Cell Sci* **123**: 4095-4106.

Martin MM, Ryan M, Kim R, Zakas AL, Fu H, Lin CM, Reinhold WC, Davis SR, Bilke S, Liu H et al. 2011. Genome-wide depletion of replication initiation events in highly transcribed regions. *Genome research* **21**: 1822-1832.

Masumoto H, Hawke D, Kobayashi R, Verreault A. 2005. A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature* **436**: 294-298.

Mattarocci S, Shyian M, Lemmens L, Damay P, Altintas Dogus M, Shi T, Bartholomew Clinton R, Thomä NH, Hardy Christopher FJ, Shore D. 2014. Rif1 Controls DNA Replication Timing in Yeast through the PP1 Phosphatase Glc7. *Cell reports* **7**: 62-69.

McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic acids research* **40**: 4288-4297.

McKay DJ, Klusza S, Penke TJR, Meers MP, Curry KP, McDaniel SL, Malek PY, Cooper SW, Tatomer DC, Lieb JD et al. 2015. Interrogating the Function of Metazoan Histones using Engineered Gene Clusters. *Developmental Cell* **32**: 373-386.

McKay Daniel J, Lieb Jason D. 2013. A Common Set of DNA Regulatory Elements Shapes Drosophila Appendages. *Developmental Cell* **27**: 306-318.

Meselson M, Stahl FW. 1958. The replication of DNA in Escherichia coli. *Proceedings of the National Academy of Sciences* **44**: 671-682.

Mesner LD, Valsakumar V, Cieślik M, Pickin R, Hamlin JL, Bekiranov S. 2013. Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. *Genome research* **23**: 1774-1788.

Milo R, Jorgensen P, Moran U, Weber G, Springer M. 2010. BioNumbers--the database of key numbers in molecular and cell biology. *Nucleic acids research* **38**: D750-D753.

Miotto B, Ji Z, Struhl K. 2016. Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. *Proceedings of the National Academy of Sciences of the United States of America* **113**: E4810-4819.

Moindrot B, Audit B, Klous P, Baker A, Thermes C, de Laat W, Bouvet P, Mongelard F, Arneodo A. 2012. 3D chromatin conformation correlates with replication timing and is conserved in resting cells. *Nucleic acids research* **40**: 9470-9481.

Munden A, Rong Z, Sun A, Gangula R, Mallal S, Nordman JT. 2018. Rif1 inhibits replication fork progression and controls DNA copy number in Drosophila. *eLife* **7**: e39140.

Newville M, Stensitzki T, Allen DB, Ingargiola A. 2014. LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python. *Zenodo* doi:10.5281/zenodo.11813.

Nishitani H, Lygerou Z, Nishimoto T, Nurse P. 2000. The Cdt1 protein is required to license DNA for replication in fission yeast. *Nature* **404**: 625-628.

Nordman JT, Kozhevnikova EN, Verrijzer CP, Pindyurin AV, Andreyeva EN, Shloma VV, Zhimulev IF, Orr-Weaver TL. 2014. DNA Copy-Number Control through Inhibition of Replication Fork Progression. *Cell reports* **9**: 841-849.

Oldach P, Nieduszynski CA. 2019. Cohesin-Mediated Genome Architecture Does Not Define DNA Replication Timing Domains. *Genes* **10**: 196.

Orr-Weaver TL. 2015. When bigger is better: the role of polyploidy in organogenesis. *Trends in genetics : TIG* **31**: 307-315.

Ozgyin L, Horvath A, Hevessy Z, Balint BL. 2019. Extensive epigenetic and transcriptomic variability between genetically identical human B-lymphoblastoid cells with implications in pharmacogenomics research. *Scientific reports* **9**: 4889.

Peace JM, Ter-Zakarian A, Aparicio OM. 2014. Rif1 Regulates Initiation Timing of Late Replication Origins throughout the *S. cerevisiae* Genome. *PloS one* **9**: e98501.

Peacock JA. 1983. Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices of the Royal Astronomical Society* **202**: 615-627.

Peng JC, Karpen GH. 2006. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nature cell biology* **9**: 25-35.

Peng JC, Karpen GH. 2009. Heterochromatin Genome Stability Requires Regulators of Histone H3 K9 Methylation. *PLoS genetics* **5**: e1000435.

Penke TJR, McKay DJ, Strahl BD, Matera AG, Duronio RJ. 2016. Direct interrogation of the role of H3K9 in metazoan heterochromatin function. *Genes & Development* **30**: 1866-1880.

Penke TJR, McKay DJ, Strahl BD, Matera AG, Duronio RJ. 2018. Functional Redundancy of Variant and Canonical Histone H3 Lysine 9 Modification in Drosophila. *Genetics* **208**: 229-244.

Petruk S, Black KL, Kovermann SK, Brock HW, Mazo A. 2013. Stepwise histone modifications are mediated by multiple enzymes that rapidly asociate with nascent DNA during replication. *Nature Communications* **4**.

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

Pfister Sophia X, Ahrabi S, Zalmas L-P, Sarkar S, Aymard F, Bachrati Csanád Z, Helleday T, Legube G, La Thangue Nicholas B, Porter Andrew CG et al. 2014. SETD2-Dependent Histone H3K36 Trimethylation Is Required for Homologous Recombination Repair and Genome Stability. *Cell reports* **7**: 2006-2018.

Pindyurin AV, Boldyreva LV, Shloma VV, Kolesnikova TD, Pokholkova GV, Andreyeva EN, Kozhevnikova EN, Ivanoschuk IG, Zarutskaya EA, Demakov SA et al. 2008. Interaction between the Drosophila heterochromatin proteins SUUR and HP1. *Journal of Cell Science* **121**: 1693-1703.

Platt EJ, Smith L, Thayer MJ. 2018. L1 retrotransposon antisense RNA within ASAR lncRNAs controls chromosome-wide replication timing. *The Journal of cell biology* **217**: 541-553.

Pokholkova GV, Koryakov DE, Pindyurin AV, Kozhevnikova EN, Belyakin SN, Andreyenkov OV, Belyaeva ES, Zhimulev IF. 2014. Tethering of SUUR and HP1 proteins results in delayed replication of euchromatic regions in Drosophila melanogaster polytene chromosomes. *Chromosoma* **124**: 209-220.

Pokholkova GV, Koryakov DE, Pindyurin AV, Kozhevnikova EN, Belyakin SN, Andreyenkov OV, Belyaeva ES, Zhimulev IF. 2015. Tethering of SUUR and HP1 proteins results in delayed replication of euchromatic regions in Drosophila melanogaster polytene chromosomes. *Chromosoma* **124**: 209-220.

Polak P, Karlic R, Koren A, Thurman R, Sandstrom R, Lawrence MS, Reynolds A, Rynes E, Vlahovicek K, Stamatoyannopoulos JA et al. 2015. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518**: 360-364.

Polo SE, Jackson SP. 2011. Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes & Development* **25**: 409-433.

Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, Vera DL, Wang Y, Hansen RS, Canfield TK et al. 2014. Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**: 402-405.

Powell SK, MacAlpine HK, Prinz JA, Li Y, Belsky JA, MacAlpine DM. 2015. Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. *The EMBO Journal* **34**: 531-543.

Price BD, D'Andrea AD. 2013. Chromatin Remodeling at DNA Double-Strand Breaks. *Cell* **152**: 1344-1354.

Prioleau M, MacAlpine DM. 2016. DNA replication origins—where do we begin? *Genes & Development* **30**: 1683-1697.

Prud'homme N, Gans M, Masson M, Terzian C, Bucheton A. 1995. Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. *Genetics* **139**: 697-711.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Quivy J-P, Gerard A, Cook AJL, Roche D, Almouzni G. 2008. The HP1-p150/CAF-1 interaction is required for pericentric heterochromatin replication and S-phase progression in mouse cells. *Nature structural & molecular biology* **15**: 972-979.

Ready DF, Hanson TE, Benzer S. 1976. Development of the Drosophila retina, a neurocrystalline lattice. *Developmental Biology* **53**: 217-240.

Reish O, Orlovski A, Mashevitz M, Sher C, Libman V, Rosenblat M, Avivi L. 2003. Modified allelic replication in lymphocytes of patients with neurofibromatosis type 1. *Cancer genetics and cytogenetics* **143**: 133-139.

Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JFX. 2009. Concerted Loading of Mcm2–7 Double Hexamers around DNA during DNA Replication Origin Licensing. *Cell* **139**: 719-730.

Rhind N, Gilbert DM. 2013. DNA Replication Timing. *Cold Spring Harbor perspectives in biology* **5**.

Rhind N, Yang SC-H, Bechhoefer J. 2010. Reconciling stochastic origin firing with defined replication timing. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **18**: 35-43.

Riddle NC, Minoda A, Kharchenko PV, Alekseyenko AA, Schwartz YB, Tolstorukov MY, Gorchakov AA, Jaffe JD, Kennedy C, Linder-Basso D et al. 2011. Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome research* **21**: 147-163.

Rivera-Mulia JC, Buckley Q, Sasaki T, Zimmerman J, Didier RA, Nazor K, Loring JF, Lian Z, Weissman S, Robins AJ et al. 2015. Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. *Genome research* **25**: 1091-1103.

Rivera-Mulia JC, Desprat R, Trevilla-Garcia C, Cornacchia D, Schwerer H, Sasaki T, Sima J, Fells T, Studer L, Lemaitre J-M et al. 2017. DNA replication timing alterations identify common markers between distinct progeroid diseases. *Proceedings of the National Academy of Sciences* **114**: E10972-E10980.

Rivera-Mulia JC, Gilbert DM. 2016. Replication timing and transcriptional control: beyond cause and effect — part III. *Current Opinion in Cell Biology* **40**: 168-178.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

Rotelli MD, Policastro RA, Bolling AM, Killion AW, Weinberg AJ, Dixon MJ, Zentner GE, Walczak CE, Lilly MA, Calvi BR. 2019. A Cyclin A-Myb-MuvB-Aurora B network

regulates the choice between mitotic cycles and polyploid endoreplication cycles. *PLoS genetics* **15**: e1008253.

Rowles A, Chong JP, Brown L, Howell M, Evan GI, Blow JJ. 1996. Interaction between the origin recognition complex and the replication licensing system in Xenopus. *Cell* **87**: 287-296.

Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, Lin MF et al. 2010. Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. *Science* **330**: 1787-1797.

Ryba T, Battaglia D, Chang BH, Shirley JW, Buckley Q, Pope BD, Devidas M, Druker BJ, Gilbert DM. 2012. Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. *Genome research* **22**: 1833-1844.

Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, Schulz TC, Robins AJ, Dalton S, Gilbert DM. 2010. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome research* **20**: 761-770.

Sasaki T, Rivera-Mulia JC, Vera D, Zimmerman J, Das S, Padget M, Nakamichi N, Chang BH, Tyner J, Druker BJ et al. 2017. Stability of patient-specific features of altered DNA replication timing in xenografts of primary human acute lymphoblastic leukemia. *Experimental Hematology* **51**: 71-82.e73.

Sato Y, Yoshizato T, Shiraishi Y, Maekawa S, Okuno Y, Kamura T, Shimamura T, Sato-Otsubo A, Nagae G, Suzuki H et al. 2013. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nature genetics* **45**: 860-867.

Schmid M, Durussel T, Laemmli UK. 2004. ChIC and ChEC; genomic mapping of chromatin proteins. *Molecular cell* **16**: 147-157.

Schwaiger M, Kohler H, Oakeley EJ, Stadler MB, Schübeler D. 2010. Heterochromatin protein 1(HP1) modulates replication timing of the *Drosophila* genome. *Genome research* **20**: 771-780.

Schwaiger M, Stadler MB, Bell O, Kohler H, Oakeley EJ, Schübeler D. 2009. Chromatin state marks cell-type- and gender-specific replication of the Drosophila genome. *Genes & Development* **23**: 589-601.

Seller CA, Cho C-Y, O'Farrell PH. 2019. Rapid embryonic cell cycles defer the establishment of heterochromatin by Eggless/SetDB1 in Drosophila. . *Genes & Development* **in press**.

Seller CA, O'Farrell PH. 2018. Rif1 prolongs the embryonic S phase at the Drosophila mid-blastula transition. *PLoS Biol* **16**: e2005687.

Seller CA, O'Farrell PH. 2018. Rif1 prolongs the embryonic S phase at the Drosophila mid-blastula transition. *PLOS Biology* **16**: e2005687.

Sher N, Bell GW, Li S, Nordman JT, Eng T, Eaton ML, MacAlpine DM, Orr-Weaver TL. 2012. Developmental control of gene copy number by repression of replication initiation and fork progression. *Genome research* **22**: 64-75.

Shermoen AW, McCleland ML, O'Farrell PH. 2010. Developmental Control of Late Replication and S Phase Length. *Current Biology* **20**: 2067-2077.

Sidler C, Li D, Wang B, Kovalchuk I, Kovalchuk O. 2014. SUV39H1 downregulation induces deheterochromatinization of satellite regions and senescence after exposure to ionizing radiation. *Frontiers in Genetics* **5**.

Siefert JC, Georgescu C, Wren JD, Koren A, Sansam CL. 2017. DNA replication timing during development anticipates transcriptional programs and parallels enhancer activation. *Genome research* **27**: 1406-1416.

Sima J, Chakraborty A, Dileep V, Michalski M, Klein KN, Holcomb NP, Turner JL, Paulsen MT, Rivera-Mulia JC, Trevilla-Garcia C et al. 2019. Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication. *Cell* **176**: 816-830.e818.

Sima J, Gilbert DM. 2014. Complex correlations: replication timing and mutational landscapes during cancer and genome evolution. *Current Opinion in Genetics & Development* **25**: 93-100.

Sims RJ, Reinberg D. 2008. Is there a code embedded in proteins that is based on post-translational modifications? *Nature reviews Molecular cell biology* **9**: 815-820.

Skene PJ, Henikoff S. 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**: e21856.

Slade D. 2018. Maneuvers on PCNA Rings during DNA Replication and Repair. *Genes* **9**: 416-443.

Smit A, Hubley R, Green P. 2013-2015. RepeatMasker Open-4.0.

Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, Allis CD, Lucchesi JC. 2000. The Drosophila MSL Complex Acetylates Histone H4 at Lysine 16, a Chromatin Modification Linked to Dosage Compensation. *Molecular and Cellular Biology* **20**: 312-318.

Smith L, Plug A, Thayer M. 2001. Delayed replication timing leads to delayed mitotic chromosome condensation and chromosomal instability of chromosome translocations. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 13300-13305.

Spradling A, Orr-Weaver T. 1987. Regulation of DNA replication during Drosophila development. *Annual review of genetics* **21**: 373-403.

Sreesankar E, Bharathi V, Mishra RK, Mishra K. 2015. Drosophila Rif1 is an essential gene and controls late developmental events by direct interaction with PP1-87B. *Sci Rep* **5**.

Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. 2009. Human mutation rate associated with DNA replication timing. *Nature genetics* **41**: 393.

Stinchcomb DT, Struhl K, Davis RW. 1979. Isolation and characterisation of a yeast chromosomal replicator. *Nature* **282**: 39-43.

Strom AR, Emelyanov AV, Mir M, Fyodorov DV, Darzacq X, Karpen GH. 2017. Phase separation drives heterochromatin domain formation. *Nature* **547**: 241-245.

Sukackaite R, Cornacchia D, Jensen MR, Mas PJ, Blackledge M, Enervald E, Duan G, Auchynnikava T, Kohn M, Hart DJ et al. 2017. Mouse Rif1 is a regulatory subunit of protein phosphatase 1 (PP1). *Scientific reports* **7**: 2119.

Sukhanova MJ, Du W. 2008. Control of cell cycle entry and exiting from the second mitotic wave in the Drosophila developing eye. *BMC Developmental Biology* **8**: 7-7.

Supek F, Lehner B. 2015. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521**: 81-84.

Sykes SM, Mellert HS, Holbert MA, Li K, Marmorstein R, Lane WS, McMahon SB. 2006. Acetylation of the p53 DNA-binding domain regulates apoptosis induction. *Mol Cell* **24**: 841-851.

Takahashi S, Miura H, Shibata T, Nagao K, Okumura K, Ogata M, Obuse C, Takebayashi S-i, Hiratani I. 2019. Genome-wide stability of the DNA replication program in single mammalian cells. *Nature genetics* **51**: 529-540.

Takebayashi S-i, Ryba T, Gilbert DM. 2012. Developmental control of replication timing defines a new breed of chromosomal domains with a novel mechanism of chromatin unfolding. *Nucleus* **3**: 500-507.

Tardat M, Brustel J, Kirsh O, Lefevbre C, Callanan M, Sardet C, Julien E. 2010. The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. *Nature cell biology* **12**: 1086-1093.

Taylor JH. 1958. The mode of chromosome duplication in Crepis capillaris. *Exp Cell Res* **15**: 350-357.

Taylor JH. 1960. Asynchronous Duplication of Chromosomes in Cultured Cells of Chinese Hamster. *The Journal of Biophysical and Biochemical Cytology* **7**: 455-463.

Team RC. 2017. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*.

Touret F, Guiguen F, Greenland T, Terzian C. 2014. In between: gypsy in Drosophila melanogaster reveals new insights into endogenous retrovirus evolution. *Viruses* **6**: 4914-4925.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**: 562.

Tuzon CT, Spektor T, Kong X, Congdon LM, Wu S, Schotta G, Yokomori K, Rice JC. 2014. Concerted Activities of Distinct H4K20 Methyltransferases at DNA Double-Strand

Breaks Regulate 53BP1 Nucleation and NHEJ-Directed Repair. *Cell reports* **8**: 430-438.

Unnikrishnan A, Gafken PR, Tsukiyama T. 2010. Dynamic changes in histone acetylation regulate origins of DNA replication. *Nature structural & molecular biology* **17**: 430-437.

Venter JC Adams MD Myers EW Li PW Mural RJ Sutton GG Smith HO Yandell M Evans CA Holt RA et al. 2001. The Sequence of the Human Genome. *Science* **291**: 1304-1351.

Visser AE, Eils R, Jauch A, Little G, Bakker PJ, Cremer T, Aten JA. 1998. Spatial distributions of early and late replicating chromatin in interphase chromosome territories. *Exp Cell Res* **243**: 398-407.

Watanabe Y, Fujiyama A, Ichiba Y, Hattori M, Yada T, Sakaki Y, Ikemura T. 2002. Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: disease-related genes in timing-switch regions. *Human molecular genetics* **11**: 13-21.

Watanabe Y, Maekawa M. 2010. Spatiotemporal regulation of DNA replication in the human genome and its association with genomic instability and disease. *Curr Med Chem* **17**: 222-233.

Watson JD, Crick FH. 1953. The structure of DNA. *Cold Spring Harbor symposia on quantitative biology* **18**: 123-131.

Woodfine K, Fiegler H, Beare DM, Collins JE, McCann OT, Young BD, Debernardi S, Mott R, Dunham I, Carter NP. 2004. Replication timing of the human genome. *Human Molecular Genetics* **13**: 191-202.

Wurtele H, Kaiser GS, Bacal J, St-Hilaire E, Lee E-H, Tsao S, Dorn J, Maddox P, Lisby M, Pasero P et al. 2012. Histone H3 Lysine 56 Acetylation and the Response to DNA Replication Fork Damage. *Molecular and Cellular Biology* **32**: 154-172.

Xiang W, Roberti MJ, Hériché J-K, Huet S, Alexander S, Ellenberg J. 2018. Correlative live and super-resolution imaging reveals the dynamic structure of replication domains. *The Journal of Cell Biology* **217**: 1973-1984.

Yamazaki S, Ishii A, Kanoh Y, Oda M, Nishito Y, Masai H. 2012. Rif1 regulates the replication timing domains on the human genome. *The EMBO Journal* **31**: 3667-3677.

Yang SC-H, Rhind N, Bechhoefer J. 2010. Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Mol Syst Biol* **6**: 404-404.

Yarosh W, Spradling AC. 2014. Incomplete replication generates somatic DNA alterations within Drosophila polytene salivary gland cells. *Genes & Development* **28**: 1840-1855.

Yeeles JTP, Deegan TD, Janska A, Early A, Diffley JFX. 2015. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**: 431-435.

Yokochi T, Poduch K, Ryba T, Lu J, Hiratani I, Tachibana M, Shinkai Y, Gilbert DM. 2009. G9a selectively represses a class of late-replicating genes at the nuclear periphery. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 19363-19368.

Yuan K, O'Farrell PH. 2016. TALE-light imaging reveals maternally guided, H3K9me2/3-independent emergence of functional heterochromatin in Drosophila embryos. *Genes & Development* doi:10.1101/gad.272237.115.

Zentner GE, Kasinathan S, Xin B, Rohs R, Henikoff S. 2015. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nature Communications* **6**: 8733.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**: R137.

Zhang Y, Malone JH, Powell SK, Periwal V, Spana E, MacAlpine DM, Oliver B. 2010. Expression in Aneuploid Drosophila S2 Cells. *PLOS Biology* **8**: e1000320.

Zhao L, Wang D, Liu Y, Chen S, Sun F. 2013. Histone acetyltransferase hMOF promotes S phase entry and tumorigenesis in lung cancer. *Cellular Signalling* **25**: 1689-1698.

Zhu X, He F, Zeng H, Ling S, Chen A, Wang Y, Yan X, Wei W, Pang Y, Cheng H et al. 2014. Identification of functional cooperative mutations of SETD2 in human acute leukemia. *Nature genetics* **46**: 287-293.

Zielke N, Edgar BA, DePamphilis ML. 2013. Endoreplication. *Cold Spring Harbor perspectives in biology* **5**: a012948.

Zielke N, Kim KJ, Tran V, Shibutani ST, Bravo MJ, Nagarajan S, van Straaten M, Woods B, von Dassow G, Rottig C et al. 2011. Control of Drosophila endocycles by E2F and CRL4(CDT2). *Nature* **480**: 123-127.