ABSTRACT

BROOKS, ASHLEY MEREDITH. DNA Replication timing in Arabidopsis thaliana. (Under the direction of Linda Hanley-Bowdoin.)

DNA replication is the duplication of the genome required for inheritance of genetic material from a parent to a daughter cell. Strict regulation of DNA replication is required to ensure that the genome is duplicated in a timely manner only once during the cell cycle. In higher eukaryotes, replication proceeds via a conserved spatio-temporal pattern referred to as the replication timing (RT) profile. Most RT profiling studies have focused on Drosophila, human and murine developmental systems. Changes in RT profiles are often associated with changes in gene expression and the epigenome. There is evidence to suggest that the DNA replication program in plants is regulated by similar mechanisms as those utilized in other higher eukaryotes. However, our ability to study the relationships between replication, transcription, and epigenetics has been hindered by the lack of genome-wide RT profiles for plants. Arabidopsis thaliana has a small, highly annotated genome, making it a good candidate for plant RT profiling. The objective of this research was to study RT in an Arabidopsis suspension cell culture. We profiled RT in light- and dark-adapted cells and compared timing changes to epigenetic and transcriptional data for both cell lines. To create RT profiles, we pulse labeled replicating DNA with the nucleoside analog 5’-Ethynl-2’-deoxyuridine (EdU) that could subsequently be conjugated to a fluorescent epitope (AF488). We then used flow cytometry to sort isolated nuclei by DNA content and EdU incorporation, obtaining three pools representing three separate portions of S-phase. DNA was isolated from each pool followed by shearing and immunoprecipitation of AF488 to isolate newly synthesized DNA from each of the S-phase fractions. Illumina paired-end sequencing of DNA from each fraction (as well as from a control sample of DNA from unlabeled, non-replicating nuclei) was used to generate RT profiles in light and dark-adapted cells. A novel algorithm was employed to segment the genome into seven RT classes, based on the predominant time of replication at each locus. Genome wide transcriptional and epigenetic enrichment were compared to the RT data. We found
that replication time in Arabidopsis is associated with transcriptional activity and relative enrichment of H3K4me3, H3K27me3 and H3K56ac in different regions of the genome. The RT pattern was highly similar between the light and dark-adapted cells. Thus far we have been unable to detect significant large-scale changes in transcription and epigenetic levels at regions that changed RT. However, regions of the pericentromere displayed a notable shift to later timing in the dark-adapted cells, suggesting a genomic plasticity in or near the centromere of Arabidopsis.
©Copyright 2016 Ashley Meredith Brooks

All Rights Reserved
DNA Replication timing in *Arabidopsis thaliana*

by
Ashley Meredith Brooks

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Functional Genomics

Raleigh, NC

2016

APPROVED BY:

_________________________________________
José M. Alonso

_________________________________________
Dahlia Nielsen

_________________________________________
Linda Hanley-Bowdoin
Co-chair of Advisory Committee

_________________________________________
William Thompson
Co-chair of Advisory Committee
I was told that I could write anything I wanted for my biography. That got me to thinking about what exactly a short biography might contain. I suppose I could write about where I was born (Pensacola Beach, FL, USA), where I’ve lived (okay - PB, FL; Fort Walton Beach, FL; New Orleans, LA; Hudson, NC; Mount Airy, NC; Greensboro, NC; Baton Rouge, LA; Raleigh, NC) and so on. Or perhaps more relevant for a biography is a summary of my life, which might include: Educational achievements? Places I’ve been? Places I wish to visit (HIMALAYAS! What region you ask? ANY REGION!)? Life lessons? Family history? It occurred to me that while such information might be useful to understand some things in the past that have shaped me, they don’t necessarily reveal much about who I really am right now or things that are in my mind. So, instead of a formal biography I’ve decided to make a list of a few things I love or like a lot, in the hope that this will better inform an interested reader about me. In no particular order:
The sound of wind chimes through an open window; Walking on gravel; Watching animals play; The smell of NC woods; Mountains, mountains, mountains; Spicy food; Halloween; Fiddle (any style); Running on trails in the woods; Yard work but not gardening; The sound of the ocean and the silence of snow; Smells - lavender, citrus, markers, paper; A good night’s sleep; Fall, Winter, Spring, Summer, in that order; Surprises; Good beer and red wine blends; Fantasy (magic, goblins, elves, kender TASSLEHOFF BURRFOOT, dragons, etc) and horror books; Scary movies; Moonlight Sonata and Für Elise.
ACKNOWLEDGEMENTS

I would first like to thank my advisers Linda Hanley-Bowdoin and Bill Thompson for their immeasurable support through the years. I am grateful to Linda and Bill for the opportunity to work with a wonderful group of people (“DNA replication group”) that includes Lorenzo Concia, Emily Wear, Emily Wheeler, George Allen, Leigh Mickelson-Young and Patrick Mulvaney. The experience has left a tremendous positive impression in my life. I would also like to thank my committee member Dahlia Nielsen for sitting down with me and taking the time to help me understand aspects of bioinformatics analysis that were relevant to my work. I am grateful to José Alonso for teaching me the fundamental aspects of functional genomics. I’m greatly thankful to members of Linda’s lab for their love, support and friendship over the years: Mary Beth Dallas, Maria Reyes, Catherine Doyle, Laura Greeley, Wei Shen and Cyprian Rajabu. Thanks to Donna Evans for funny conversations and exhausting jogs. Finally I would like to give a huge THANK YOU to Jared Smith for his friendship, free-time entertaining, laughing, listening, cat-wrangling, movie-watching, yard work bossing, made-up karate demonstrations, traveling, interesting conversations and mostly for loving me.
# TABLE OF CONTENTS

LIST OF TABLES .............................................................................................................................................. vi

LIST OF FIGURES ............................................................................................................................................... vii

Chapter I – DNA Replication timing in higher eukaryotes ................................................................. 1
Introduction ..................................................................................................................................................... 1
A condensed history of study.......................................................................................................................... 2
The current state ............................................................................................................................................. 11
A missing piece of the puzzle .......................................................................................................................... 13
References ....................................................................................................................................................... 17

Chapter II – Transcriptional and epigenetic shifts during dark adaptation of Arabidopsis ....................................................................................................................................................... 22
Introduction ..................................................................................................................................................... 22
Results and Discussion ..................................................................................................................................... 25
Conclusions ........................................................................................................................................................ 30
Materials and Methods .................................................................................................................................... 31
References ....................................................................................................................................................... 35

Chapter III – Epigenetic state and transcription are related to replication timing in Arabidopsis ....................................................................................................................................................... 46
Introduction ..................................................................................................................................................... 46
Results and Discussion ..................................................................................................................................... 49
Conclusions ........................................................................................................................................................ 56
Materials and Methods .................................................................................................................................... 56
References ....................................................................................................................................................... 63

Chapter IV – DNA Replication timing displays plasticity in the pericentromere during dark adaptation of Arabidopsis ....................................................................................................................................................... 78
Introduction ..................................................................................................................................................... 78
Results ............................................................................................................................................................. 80
Discussion .......................................................................................................................................................... 85
Materials and Methods .................................................................................................................................... 87
References ....................................................................................................................................................... 94

Chapter V – Future Prospectives .................................................................................................................. 106
References ....................................................................................................................................................... 111
LIST OF TABLES

TRANSCRIPTIONAL AND EPIGENETIC SHIFTS DURING DARK ADAPTATION OF ARABIDOPSIS

Table 1. Mappable read counts from Illumina paired-end data ........................................ 42
Table 2. Enriched MapMan Category terms with corrected P-value ≤ 0.01 of
Genes up-regulated and down-regulated in dark............................................................... 43
Table 3. Overlap of MACS2 peaks and genes................................................................. 44
Table 4. Comparison of MACS2 peaks calling between light and dark......................... 45

EPIGENETIC STATE AND TRANSCRIPTION ARE RELATED TO REPLICATION TIMING IN ARABIDOPSIS

Table 1. Mappable read counts from Illumina paired-end data ........................................ 75
Table 2. Description of the possible epigenetic signatures for 1-kb windows across the genome. The “-” and “+” in each column indicates a signature defined as depleted or enriched, respectively, for that epigenetic ........................................ 76
Table 3. Chi-square table comparing the distribution of RT across the various epigenetic signatures. The bold font indicates that a given RT class and epigenetic signature occurred together more often than expected by chance .... 77

DNA REPLICATION TIMING DISPLAYS PLASTICITY IN THE PERICENTROMERE DURING DARK ADAPTATION OF ARABIDOPSIS........................................ 78

Table 1. Mappable read counts from Illumina paired-end data ........................................ 104
Table 2. Potential RAT classifications. “Earlier-to-later” and “Later-to-earlier” indicates the direction of RT change from light to dark. A “Half step” refers to a change from one time class to an adjacent time class when RT is considered in a linear manner (i.e. E <-> EM <-> M <-> ML <-> L). A “Full step” is any change between timing classes that are not adjacent............. 105
LIST OF FIGURES

TRANSCRIPTIONAL AND EPIGENETIC SHIFTS DURING DARK ADAPTATION OF ARABIDOPSIS

Figure 1. Phenotypic comparison of light and dark grown cells. Cells were considered dark-adapted after two weeks in constant dark.............................................. 37

Figure 2. Mean H3K4me3 signal (a) and H3K27me3 signal (b) across genes +/− 1-kb in light and dark samples. The grey dotted line indicates IP to control ratio signal = 1 to delineate enrichment above or below background............. 38

Figure 3. Histograms displaying overlap between differentially expressed genes and H3K4me3 or H3K27me3 peaks. Histogram bars are color coded as follows: green = count of differentially expressed genes that contained an epigenetic peak only the light samples, purple = count of differentially expressed genes that contained an epigenetic peak only the dark samples, grey = count of differentially expressed genes that contained an epigenetic peak in light and dark, white = count of differentially expressed genes that did not contain an epigenetic peak in light or dark......................................................... 39

Figure 4. IGV browser shots illustrating relative RNA-Seq, H3K4me3 and H3K27me3 fold-enrichment tracks for light and dark. The relevant down-regulated (a) and up-regulated (b) gene regions are surrounded by a dotted blue line. A solid black line beneath a peak indicates a region of epigenetic enrichment as determined by the MACS2 peak caller................................. 40

Figure 5. Mean H3K4me3 signal (a,b) and H3K27me3 (c,d) across genes +/− 1-kb that are up-regulated (a,c) or down-regulated (b,d) in the dark........................................ 41

EPIGENETIC STATE AND TRANSCRIPTION ARE RELATED TO REPLICATION TIMING IN ARABIDOPSIS

Figure 1. Representative FACS profile of Arabidopsis suspension cell nuclei. Sort gates are indicated by the red box. The early, mid and late S and G1 nuclei fractions were collected ........................................................................................................ 67

Figure 2. Early, mid and late S-phase sequencing read pile-up normalized against the G1 reference. Merged biorep data for all five Arabidopsis chromosomes are shown. The small black boxes mark the centromere for each chromosome.......................................................................................................................... 68

Figure 3. Chromosome III IGV screen shot illustrating RT tracks and the corresponding segmentation track......................................................................................................................... 69

Figure 4. General features of Arabidopsis RT. Panel (a) displays the RT pattern for each of the five chromosomes. The distribution of RT classes across the Genome is represented in the pie chart (b). Panel (c) displays the size distribution of RT classes. The boxes in the plots represent the 25-75th inner
quartile range (IQR) with the median indicated by a solid black line. Lower and upper whiskers correspond to $25^{th}$ quartile – $1.5 \times \text{IQR}$ and $75^{th}$ quartile + $1.5 \times \text{IQR}$, respectively ................................................................. 70

Figure 5. Chromosome II RT and segmentation tracks. The zoom shows an atypical region of late replication (red) in the short of chromosome II corresponding to the NOR .................................................................................................................. 71

Figure 6. Average epigenetic FE level across RT classes. The dotted black line indicates IP:control ratio = 1 to delineate enrichment or depletion of epigenetic signal.................................................................................................... 72

Figure 7. Distribution of RT classes across epigenetic states. The control panel represents the distribution of RT classes across the genome......................................................... 73

Figure 8. Bar plot of the mean level of transcription across RT classes. Groups marked with the same letter are not significantly different............................................. 74

DNA REPLICATION TIMING DISPLAYS PLASTICITY IN THE PERICENTROMERE DURING DARK ADAPTATION OF ARABIDOPSIS

Figure 1. IGV tracks displaying chromosome III RT and segmentation profiles....... 97
Figure 2. General features of RT distribution for light and dark. Pie chart displaying the distribution of RT classes across the genome for light (a) and dark (b) grown cells. Box and whisker plot of distribution of RT segment sizes for light (c) and dark (d). Boxes in the plots represent the $25$-$75^{th}$ inner quartile range (IQR) with the median indicated by a solid black line. Lower and upper whiskers correspond to $25^{th}$ quartile – $1.5 \times \text{IQR}$ and $75^{th}$ quartile + $1.5 \times \text{IQR}$, respectively .................................................................................................................. 98

Figure 3. IGV image of chromosome III RT segmentation for light and dark samples with corresponding RAT below. The RAT track is color coded to indicate direction and degree of RT changes as follows: Earlier-to-later, half step = light blue; Earlier-to_later, full step = dark blue; Later-to-earlier, half step = light red; Later-to_earlier, full step = dark red................................................................. 99

Figure 4. Size distribution of RATs. Boxes in the plots represent the $25$-$75^{th}$ inner quartile range (IQR) with the median indicated by a solid black line. Lower and upper whiskers correspond to $25^{th}$ quartile – $1.5 \times \text{IQR}$ and $75^{th}$ quartile + $1.5 \times \text{IQR}$, respectively .................................................................................................................. 100

Figure 5. Scatter plots of epigenetic and transcriptional data at RATS. The light ($x$-axis) and dark ($y$-axis) FE values for epigenetic marks and transcriptional data, at each 1-kb window that comprises a RAT, were plotted................................. 101

Figure 6. Distribution of epigenetic and transcriptional differences at RATS with corresponding box and whiskers plots. Boxes in the plots represent the $25$-$75^{th}$ Inner quartile range (IQR) with the median indicated by a solid line. Lower and upper whiskers correspond to $25^{th}$ quartile – $1.5 \times \text{IQR}$ and $75^{th}$ quartile + $1.5 \times \text{IQR}$, respectively. Outliers are indicated by black dots that lie outside of the whiskers.................................................................................................................. 102
Figure 7. IGV image of Light and Dark RT segmentation with corresponding outlier RAT density track. The upper panel shows distribution of RATs across the chromosomes. The zoom displays a close view of chromosome III. Each tick mark in the outlier track represents a RAT.
Chapter I

DNA Replication timing in higher eukaryotes

Introduction

DNA replication refers to the duplication of the genome that is required for the inheritance of genetic material from a parent to a daughter cell. To discuss replication, it is useful to provide a reminder of the cell cycle. From the most simple view, the cell cycle can be divided into Interphase, which includes DNA replication, and Mitosis, the phase in which the cell divides. Since replication is important to this discussion, I will describe Interphase in slightly greater detail. Interphase is divided into: Gap1 (G1) phase during which the cell expands and replication proteins are synthesized, Synthesis (S) phase which is the phase referring specifically to DNA replication, and Gap2 (G2) phase that includes preparation for cell division.

In higher eukaryotes DNA replication is initiated during G1-phase with assembly of the origin recognition complex (ORC) at genomic sites referred to as “origins” (Chagin, et al 2010). Binding of ORC to the DNA strand leads to formation of the pre-replication complex (pre-RC) via recruitment of the “licensing factors” Cdc6, Cdt1, MCM9 and MCM2-7. The origin is considered “licensed” once the pre-RC is formed (Leonard and Méchali 2013). The pre-RC then recruits Cdc45, RPA and DNA polymerases to form the pre-initiation complex (pre-IC). The formation of the pre-IC coincides with initiation of replication, otherwise referred to as “origin firing”. Once replication has initiated at the origin, it continues outward in a bidirectional manner until it reaches a termination point. The region of DNA that is replicated from the origin to it termination points is referred to as a replicon. Eukaryotic chromosomes contain many replicons, that range between 30 and 450 kb in size in mammals (Zink 2006). The use of multiple replicons is necessary for replication to be completed within a limited time frame before cell division occurs (Chagin, et al 2010; Leman and Noguchi 2013).
While licensing of all potential origins is believed to occur simultaneously during G1, origin firing occurs at various times during S-phase, and some licensed origins never fire. Potential origins that do not fire are eventually replicated passively and are incorporated into a replicon spreading from another nearby origin. Thus, the time at which a genomic region that is not an origin replicates depends upon its distance from the nearest origin and the time during S-phase that the origin fires (Schwaiger, et al. 2009). Origins in close proximity tend to fire at the same time, and adjacent groups of replicons or “replication foci” are replicated at or near the same time. The temporal coordination of replication foci results in a distinctive replication timing (RT) profile across the genome. RT is the focus of this dissertation. Specifically we seek to understand how RT is regulated in plants. I will use the current chapter to provide background information, beginning with a general description of RT and how that knowledge was obtained. I will then discuss more current information about RT and the status of our knowledge of plant RT before the experiments described in the subsequent chapters.

**A condensed history of study**

Beginning in the mid-20th century, pulse labeling of nucleic acids was introduced and is now one of the most important techniques to study DNA replication. In brief, pulse labeling refers to the incorporation of a modified nucleoside analog into a growing nascent DNA strand for a short period of time. Depending on the specific modification (e.g. radioactive, fluorescent, epitope tag) on the nucleoside, the newly replicated DNA can be visualized or isolated for various downstream analyses (Taylor 1957; Wear, et al. 2016).

Early cytological studies involving pulse labeling of DNA with tritium labeled thymidine followed by autoradiographic visualization revealed the existence of regions of DNA that replicate at the same time during S-phase (Taylor 1957). Similar experiments also revealed that the clusters of DNA localize to specific regions of the nucleus during S-
phase. These early experiments laid a framework that eventually led to our current understanding that replication is regulated in a spatio-temporal manner during S-phase. Specific regions of the genome replicate at a distinct time in S-phase (e.g. early) while other regions replicate late. Chromatin that replicates at specific times during S-phase localizes to distinct nuclear spaces depending upon cell type and species. For example, early replicating regions of the mammalian genome tend to localize to the nuclear interior (Visser, et al. 1998; Zink et al. 2006). In contrast, early replication in maize is distributed throughout the nucleoplasm. In mammals, middle S-phase replication occurs in the perinuclear and perinucleolar space while in plants middle S-phase replication co-localizes with early replication in the nucleoplasm (Bass et al. 2015).

The spatio-temporal coordination of DNA replication is related to structural changes in DNA that coincide with its localization to a specific nuclear compartment during S-phase. These changes are related to the degree of chromosomal condensation that ranges from a relaxed formation to a highly condensed state. To facilitate packaging and storage of large amounts of genetic material in the nucleus, DNA is complexed with proteins that enable the DNA molecule to fold and bend. The complex of DNA and protein is referred to as chromatin. The primary building block of chromatin is a core nucleosome that consists of ~146 base pairs of DNA wrapped twice around an octamer of histones (Luger, et al. 1997). The histone octamer is comprised of two dimers of H2A and H2B and a tetramer of H3 and H4. The amino-terminal tail of each histone and the carboxy-terminal tails of H2A and H2B protrude from the octamer. The residues that comprise these tails and those that are near the nucleosome surface play a key role in the regulation of chromatin dynamics. The histone linker H1 is bound to DNA outside of the core nucleosome. The strength of interactions between the core nucleosome, H1 and their associated DNA can be weakened or strengthened by a variety of mechanisms. Weaker interactions allow a relatively open relaxed chromatin state termed “euchromatin,” while stronger interactions result in a more highly condensed state.
termed “heterochromatin”. These types of chromatin can be distinguished with chemical staining. Microscopy studies revealed that heterochromatin is late replicating while euchromatin replicates earlier and that these different types of chromatin occupy distinct locations in the nucleus during S-phase (Ahmad and Henikoff 2001).

Cytological research has focused on understanding replication in nuclear space and time. In contrast, there is a large body of research surrounding the role of nuclear dynamics in transcription. Scientists have long known that the level of transcriptional activity differs across the genome and that the chromatin state is related to these differences. For example, euchromatin is associated with active gene expression while heterochromatin is associated with repressed transcription (Wolffe and Pruss 1996; Gregory and Horz 1998; Wakimoto 1998). The relative position of a gene in euchromatin or heterochromatin may activate or repress expression of the gene (Henikoff 1992; Csink and Henikoff 1996). Regulatory elements can affect gene positioning relative to euchromatin and heterochromatin. For example, an earlier experiment determined that enhancers prevent gene silencing by interfering with the localization of the corresponding gene to heterochromatin (Francastel, et al. 1999). Enhancer mutations that lead to increased silencing also prevent the regulated gene from relocating away from heterochromatin. This suggests a dual role for enhancers in transcription and gene positioning in the nucleus. Collectively, these studies provide insight into how the nucleus can accomplish multiple cellular processes at once in an efficient manner. Compartamentalization of euchromatin with processes that require access to DNA (i.e. transcription and replication) may reduce interference from heterochromatic spreading and interactions with factors that could inhibit transcription and replication.

The first direct links between RT and transcription were based on the analysis of a few genes and often related to genetic changes that occur during development or cell differentiation (Simon, et al. 2001). An example is illustrated by studies of
differentiation of human embryonic stem (ES) cells into neural cells. Genes involved in maintaining the ES cell state have higher levels of expression and replicate early in ES cells. In neural cells, these genes are repressed and switch to a later replication timing (Perry, et al. 2004). However, the field of RT changed dramatically with the application of genome-wide approaches that can identify patterns that are not observable at a small scale. The application of microarray and DNA sequencing technology enabled the mapping of RT to the whole genome in Drosophila and mammals for the first time (Schubeler, et al. 2002; Woodfine, et al. 2004; Hiratani, et al. 2008; Hansen, et al. 2010).

Although slight modifications exist, two basic data analysis approaches have been applied to mapping RT with microarrays. One approach, first used in Drosophila and murine cells, relies upon pulse labeling DNA with a nucleotide analog, such as BrdU, in an asynchronous population of cells. Cells are then sorted into S-phase fractions based on total genomic content using fluorescence activated cell sorting (FACS). DNA labeled during early or late S-phase is immunoprecipitated using an antibody to the nucleotide analog and sequence enrichment is analyzed using microarrays. The RT of each genomic region represented on the microarray is calculated as a ratio of early to late signal. A ratio >1 or <1 indicates an early replicating region or a late replicating region, respectively (Schubeler, et al. 2002; Hiratani, et al. 2008). A second microarray approach generated the first RT map in humans (Woodfine, et al. 2004). This method, which is derived from copy number analysis, involves labeling double-stranded DNA with Hoechst 33258 in an asynchronous cell population. The cells are sorted into a replicating S-phase population and a non-replicating G1 control using FACS. DNA from each population is isolated and hybridized to a microarray probes. The signal ratio of S-phase to G1 is determined for sequences represented on the array. A ratio of 2:1 for a sequence indicates that there are two copies of that sequence represented in the S-phase population and is early replicating. In contrast, a ratio of 1:1 indicates that the sequence has not yet been replicated and is late replicating. The plot of the ratio, which can vary between 2:1 and 1:1, is used to generate the RT profile (Woodfine, et al. 2004).
Despite differences in techniques, a common thread between these two microarray approaches is the use of ratio of signals derived from early and late phases of the cell cycle to determine RT. Although the experimental approaches described above rely upon FACS to obtain synchronized populations of cells, synchronization can be accomplished through other techniques. For example, chemicals such as aphidicolin, hydroxyurea and nocodazole can be used to block the progression of the cell cycle at various stages (Pagano, et al 1992; Cooper and Shedden 2003). Other methods of synchronization include mechanical means (eg “shake off”) or selection of cells based on properties such as size or age. However, it is possible that disruption of the cell cycle with these synchronizing methods can introduce disruptions or artifacts in the analysis of replication timing and transcription (Cooper and Shedden 2003).

The application of sequencing to generate genome-wide RT profiles, termed “Repli-Seq”, was recently applied to human cells (Hansen, et al. 2010; Pope, et al. 2014). Similar to the microarray approach described earlier, cells are pulse labeled with a nucleotide analog such as BrdU and sorted based on DNA content into various S-phase fractions, as well as G1 and/or G2-phase. DNA is extracted (G1) and immunoprecipitated (S-phase and G2), sequenced and mapped to the genome. The read density of each S-phase fraction is normalized to a background control that is determined by the experimenter.

Genome-wide analysis allowed RT profiles to be compared to genomic features such as gene density, gene expression and transposable elements. In general, early replication is associated with regions of high gene density and higher levels of gene expression. Conversely, late replication is associated with lower gene density, decreased gene expression, and a greater density of transposons (Schubeler, et al. 2002; Hansen, et al. 2010). Furthermore, comparison of RT profiles in different cell types revealed remarkable developmental plasticity in the mammalian and Drosophila RT programs. Genomic domains that switch replication time (RDs) range from ~400-800 kb in size.
and include 20-50% of the genome (Hiratani, et al. 2008; Hansen, et al. 2010). Genomic regions that do not switch RT are referred to as “constant timing regions” (CTRs). Interestingly, CTRs and regions that switch RT have different characteristics. For example, gene density in early human CTRs is higher than the genome average while the gene density in late CTRs is typically less than the genome average (Hansen, et al. 2010). Regions that switch RT have a lower gene density compared to the genome average, more like the late CTRs (Hansen, et al. 2010).

In addition to analysis of genomic features, genome wide RT profiles allowed comparison of dynamic features such as transcriptional activity and epigenetic enrichment. Early replication is associated with high levels of transcription while late replicating regions are associated with repressed transcription (Schubeler, et al. 2002; MacAlpine, et al. 2004). These findings supported previous conclusions at the single gene level. However, there are exceptions to the pattern. For example, a study in Drosophila found that despite an association between RT and transcription, up to 30% of early replicating genes are not expressed and 30% of late replicating genes are expressed (Schubeler, et al. 2002). The conditional association is also evident in comparisons of replication time with transcription levels in different cell types. Microarray analysis of different Drosophila cell types revealed that only ~50% of the regions that switch timing displayed different levels of gene expression (Schwaiger, et al. 2009). Although RT and transcriptional activity seem to be associated, the lack of an absolute correlation indicates that their relatedness might be attributed to other underlying cellular dynamics.

A fascinating aspect of DNA regulation is epigenetics. Epigenetic factors are widely accepted as key players in the regulation of DNA RT. Broadly defined, epigenetics refers to transient, chemical modifications to the DNA strand itself, as well as histone variants and post-translational modifications to histone proteins (Berger, et al. 2009). Epigenetic markers impact many genetic processes, including, but not necessarily
limited to, transcription, chromatin formation and DNA replication. Epigenetic modifications to DNA include various methylation patterns that are sequence dependent. Histone variants replace canonical histones during the cell cycle to modulate chromatin function. The most highly studied post-translational histone modifications (HM) are acetylation and methylation, but phosphorylation, ubiquination and sumoylation are increasingly of interest as their potential roles in the cell are revealed. Collectively, the various modifications to DNA and histones are referred to as the epigenome. The existence of an association between the epigenome and RT is undeniable but our knowledge of how these processes are co-regulated is still in its infancy.

For this discussion, I will focus on post-translational modifications to histones (HM). For the cell, it is likely that the regulatory roles of histone variants and epigenetic modifications to DNA are equally important as histone modifications. However, histone variants and epigenetic DNA modifications are outside of the scope of this dissertation. To learn more about them or their role in RT, the reader is referred to Stroud, et al. (2012) or Zhou, et al. (2015) for histone variants and Du, et al. (2015) or Schubeler (2015) for DNA methylation.

HMs are involved in regulating gene expression, acting in either an activating or repressive manner. A defining characteristic of HMs is that they are transient, meaning that they can be deposited and removed as required for transcriptional control. This transience is significant because it implies a mechanism that the cell can utilize to fine tune cellular responses to developmental and environmental cues. Furthermore, the reversibility of HM imparts a certain degree of phenotypic plasticity that is required for reproduction or survival. For example, plants are vulnerable to unpredictable environmental changes from which they cannot escape due to immobility. Thus, when confronted with a changing environment, they must be able to respond immediately to the rapid onset of external challenges (e.g. predatory wounding) or they must adapt to
environmental conditions that are longer lasting (e.g. canopy shading). Epigenetic mechanisms are implicated in a number transcriptional plant responses to external factors such as light, water, salinity and temperature (Charron, et al. 2009; Kim, et al. 2010; Zong, et al. 2013; Liu, et al. 2014). Similar to environmentally triggered cellular responses, development requires precise activation and repression of developmental stage-specific genes.

An intriguing observation, first reported by Bernstein et al (2006), is the existence of “bivalent domains,” where the co-occurrence of two epigenetic marks of antagonistic function are present at the same genomic location (Bernstein et al 2006). Specifically, they found that H3K4me3 and H3K27me3 marked promoters of developmentally regulated genes that have little or no expression in embryonic stem cells (ESCs). In differentiated cells, developmental genes associated with H3K4me3 alone had much higher expression than those marked with H3K4me3 and H3K27me3. Furthermore, expression levels of genes marked with H3K27me3 alone in differentiated cells were similar to genes in ESCs that were bivalently marked, suggesting that H3K27me3 is epistatic to H3K4me3. They proposed that bivalency maintains the developmental genes in a transcriptionally silent state that is poised for expression as differentiation occurs (Bernstein et al 2006). To date, the H3K4me3/H3K27me3 bivalent regulation has been described in mammals, plants and most recently in Drosophila, indicating that it is an important conserved regulatory feature (Bernstein, et al. 2006; Brusslan, et al. 2012; Schertel, et al. 2015). The study of bivalency has been extended to include, not only combinations of two marks, but also combinations of multiple HMs that define very specific chromatin features. We will return to the discussion of the combinatorial activity of HM in the context of RT later.

Histone modifications are believed to impact transcription by altering chromatin structure. This is accomplished by directly affecting histone-DNA interactions or by recruiting chromatin modifying effector proteins (Voigt and Reinberg 2011). Depending
upon the modification, DNA-histone interactions can be relaxed or strengthened such that chromatin will adopt either a euchromatic or heterochromatic state. Euchromatin is often associated with acetylated lysine residues. Acetylation interferes with electrostatic interactions between the nucleosome and DNA, resulting in a relaxation of chromatin. An important example is acetylation of lysine 56 on H3 (H3K56ac). The lysine-56 residue resides on the surface of the nucleosome at the site of DNA contact and the presence of an acetyl group increases DNA breathing (i.e. openness) (Ozdemir, et al. 2005; Neumann, et al. 2009). A nucleosome reconstitution study focusing on DNA-H3K56ac interaction kinetics reported the effect of this HM on DNA structure was localized and had very little impact on large-scale chromatin compaction (Neumann, et al. 2009). In contrast to euchromatic HMs, modifications such as tri-methylation of H3K27 (H3K27me3) and di-methylation of H3K9 (H3K9me3) are associated with facultative and constitutive heterochromatin, respectively. There are many HMs that impact chromatin structure and previously unidentified modifications are still being discovered. It is likely that their roles, either singularly or in combination with other HMs, will continue to help elucidate more details relating to chromatin dynamics that impact RT.

Recently, a study examined the power of multiple genomic features, including 28 HMs, to predict genome-wide RT in Drosophila (Comoglio and Paro 2014). Using data from previous studies, they applied linear regression modeling to calculate the predictive power of individual and combinatorial features to determine RT across the genome in a single cell type. They found that enrichment levels of only five HMs accurately predicted RT when they are analyzed alone. For example, H4K5ac and H4K20me1 levels were predictive for early and late RT, respectively. Further analysis revealed that H3K36me1, H4K8ac and ubiquinated histone 2B (H2Ub), as individual marks, were highly predictive for early replication. Interestingly, however, the combination of H3K36me1, H3K79me1 and H2Ub was highly predictive for late replication. This indicates that the
power for a HM to predict RT may differ depending upon the context (Comoglio and Paro 2014).

The power of a combination of HMs to predict differential RT between two cell types was also assessed in Drosophila (Comoglio and Paro 2014). Several marks, both singular and in combinations were identified to predict RT. The H3K18ac and H3K79me1 marks (alone), and H3K36me1 in combination with H3K27me3, H3K4me1, and H3K36me3, were highly significant predictors for early RT. In contrast, H3K9me2, alone, and in combination with H3K4me1, accurately predicted late RT. Again we see a different RT predication (i.e. H3K4me1) depending upon the combinatorial context. In summary, this study provides evidence that the combinatorial activity of HMs impacts RT.

In our discussion so far, we have seen multiple lines of evidence suggesting strong associations between RT, transcription and epigenetic features of chromatin. We began with early cytological studies that hinted at a spatio-temporal association between RT, transcription and chromatin structure. Moving into the world of genomic analysis, we discussed connections between RT and epigenetic HMs that play a key role in the regulation of gene expression. In the remainder of this chapter, we will discuss some of the most recent work relating to RT as well as a few gaps in that field of research.

**The current state**

A fundamental characteristic of chromatin that underlies its sub-nuclear arrangement is interactions between distal and neighboring regions of chromosomes. These interactions can have functional significance and their disruption can have detrimental consequences for the cell. Many different types of chromosomal interactions exist and a few examples are genes and their regulatory elements, heterochromatin and heterochromatic HMs (i.e. H3K27me3 and H3K9me2), and between replicating foci (Nakamura, et al. 1986; Zhao, et al. 2006; Feng, et al. 2014; Rivera-Mulia and Gilbert 2016). The method of chromosome conformation capture (3C) and its derivatives were
developed to study chromosomal interactions (Dekker, et al. 2002; Zhao, et al. 2006; Lieberman-Aiden, et al. 2009). The 3C derivative “Hi-C” combines the capture of interacting DNA regions and sequencing to identify all inter- and intra- chromosomal interactions (Lieberman-Aiden, et al. 2009). Because of the identification of numerous chromosomal interactions and the decreasing cost of DNA sequencing, Hi-C has become the most widely applied 3C method for studies involving chromosomal dynamics (Sati and Cavalli 2016; Rowley and Corces 2016), and great strides have been made in this area. For example, Hi-C experiments in human cell lines revealed interactions between euchromatic, gene-dense regions, as well as strong interactions between heterochromatic, gene-poor regions (Lieberman-Aiden, et al. 2009). Furthermore, euchromatic regions are enriched for the active HM H3K36me3 and heterochromatic regions are enriched for the repressive HM H3K27me3. Thus, chromosome interaction data confirm the genome wide co-localization of specific “active” or “repressive” features that were first suggested by earlier cytological observations.

Comparison of human Hi-C interaction maps with RT profiles revealed strong similarity in their patterns. Interacting heterochromatin in the Hi-C data aligned with late replicating regions in RT experiment while euchromatin aligned with early replicating regions (Lieberman-Aiden, et al. 2009; Ryba, et al. 2010). A high correlation between these two data sets was observed for all human autosomal chromosomes. Strikingly, the genome-wide correlation was higher than correlations calculated between RT and all other genomic features (Ryba, et al. 2010).

A separate Hi-C analysis in human and murine cells revealed the existence of interactions between large chromosomal domains referred to as “topologically associated domains” (TADs) (Dixon, et al. 2012). Boundaries of TADs (i.e. regions that flank TADs) have distinct characteristics that link to transcriptional and epigenetic features. For example, TAD boundaries are associated with insulator sequences that regulate transcription and act as barriers to the spread of heterochromatin. Boundaries
are also associated with the active HMs H3K4me3 and H3K36me3 HMs. Boundaries are depleted for the heterochromatic HM H3K9me3, while TADs are enriched for H3K9me3. Together these results suggest that TADs are inactive, heterochromatic regions that are bordered by active euchromatin (Dixon, et al. 2012). Comparison of the Hi-C data and RT profiles showed a near perfect alignment of TAD boundaries with regions that change RT (Dixon, et al. 2012). The remarkable overlap of RT profiles and Hi-C data with accompanying transcriptional and epigenetic information provides strong support that RT is regulated by mechanisms that relate to the organization of chromatin in the nucleus. Furthermore, the efficient co-regulation of multiple cellular processes might be enabled by chromatin organization strategies, such as folding and looping, that allow chromosomal interactions within sub-nuclear territories.

It is interesting that housekeeping genes are enriched at TAD boundaries while tissue specific genes were enriched in TADs. One functional consequence of this arrangement could be that housekeeping genes are more readily accessible for continual transcription in a euchromatic environment. However, developmental specific genes that are in heterochromatic TADs are shielded from aberrant transcription when their expression is not required. Based on this distribution, an interesting possibility is that housekeeping genes, which are dispersed throughout the genome, are in close proximity to one another in 3D space and exist in a constant euchromatic environment with very little heterochromatin formation.

A missing piece of the puzzle

From the earliest cytological work to current genome-wide profiling studies, the study of RT in higher eukaryotes has greatly advanced. However, there is relatively little current published information about plant DNA RT. One reason for this is that plant genomes undergo polyploidy events that result in gene duplication. Gene homologs diverge over time and functional consequences are not fully understood (Roulin, et al. 2013). DNA replication proteins are conserved between plants and mammals. However,
the existence of multiple homologs encoding the replication proteins suggests that they may have unique functions in plants (Vandepoele, et al. 2002; Shultz, et al. 2007). Much of the replication timing work in plants has focused on characteristics such as replicon size, rate of replication and S-phase duration using various DNA labeling and detection procedures (Taylor et al. 1957; Van’t Hof 1976; Van’t Hof et al. 1978; Wear et al 2016). It is interesting to note that the semi-conservative nature of replication was first identified in the plant Vicia faba, the English broad bean (Taylor et al. 1957). Despite this early progress, however, high-resolution genome scale studies of replication timing have not progressed in plants nearly as far as they have in human, mouse and Drosophila.

The model plant Arabidopsis, which has a small genome with a relatively low amount of repeated sequences is an excellent system in which to study plant replication timing. The Arabidopsis genome has been very well annotated, facilitating functional studies that are limited in other plant species. DNA fiber autoradiatography revealed that replication of the Arabidopsis genome is biphasic based on the finding that two groups replicons exist and fire either early or late (Van’t Hof et al. 1978). A more recent microarray study profiled RT of Arabidopsis chromosome IV (Lee, et al. 2010). Cells were pulse labeled for one hour with BrdU, and fixed nuclei were sorted by FACS based on total DNA content. The sorted populations consisted of Early S/G1, Mid S and Late S/G2 nuclei. An anti-BrdU antibody was used to isolate the replicating sequences from each population. A segmentation algorithm was applied to segment the genome into seven RT classes. The authors reported that chromosome IV replicates in a biphasic manner, confirming the earlier DNA autoradiography work. Specifically they found that euchromatic regions replicate in E and M S-phase while heterochromatin replicates in L S-phase. Early replicating regions were associated with the euchromatic HM H3K56ac, demonstrating for the first in plants that RT is related to chromatin state. The HM H3K9me2 is enriched in the late replicating pericentromeric region of the long arm of chromosome VI, suggesting that this region is comprised of constitutive
heterochromatin. The chromosome IV data uncovered potential similarities in RT and genomic features between plant and other organisms that make Arabidopsis an appealing candidate for genome wide RT profiling.

A recent study identified ~1500 putative replication origins in Arabidopsis (Costas, et al. 2011). The putative origins were enriched at the 5’ end of genes. There was a weak correlation with expression levels but genes that were highly expressed contained more origins that genes that had low expression values. Most of the origins were associated with a high GC content and were depleted for heterochromatin associated CG methylation. Earlier replicating regions were previously shown to be depleted of CG methylation in Arabidopsis (Lee, et al. 2010). A notable level of the transcriptional histone variant H2A.Z was detected near origins. H2A.Z disrupts nucleosome stability (Jarillo and Pineiro 2015). Together these data suggest that similar mechanisms might underly replication origin function and chromatin organization in Arabidopsis.

There is other evidence that some aspects of chromatin dynamics in Arabidopsis are similar to other eukaryotes. Fluorescence in-situ hydridization (FISH) revealed that heterochromatic regions interact to form domains referred to as chromocenters (CC) (Fransz, et al. 2002). Several euchromatic loops, ranging from 0.2 – 2 Mbp, extended from the chromocenters. They found that the CCs and corresponding loops occupied distinct compartments in the nucleus, suggesting a relationship between chromatin state and nuclear location. Similar to previous results in mammals, Hi-C analysis of Arabidopsis chromatin revealed that the strongest interactions were between heterochromatic pericentromeres that most likely represent the CCs characterized in earlier cytological work. The long-range interactions between heterochromatic regions throughout the genome were correlated with heterochromatic H3K27me3 and H3K9me2. Moreover, when epigenetic pathways were mutated, there were corresponding perturbations in the interactions, indicating that the observed associations are more than coincidental (Feng, et al. 2014).
Despite some similarities, there are notable differences between Arabidopsis and other eukaryotes. One fundamental difference is that mammalian genomes contain a high number of local interactions between adjacent, large (mega-base sized) domains (Lieberman-Aiden, et al. 2009; Dixon, et al. 2012). These interacting domains are conserved between human and mouse cells (Ryba, et al. 2010). In contrast, Hi-C analysis of Arabidopsis revealed that interacting regions are dispersed across the genome and much smaller (Feng, et al. 2014). It is important to keep in mind that the Arabidopsis genome is relatively small and gene dense compared to many other plant species. Therefore, we cannot conclude that the observed interaction pattern for Arabidopsis is a characteristic of plants in general and, instead, may be more related to the structure of the Arabidopsis genome specifically. It is likely that this information will be better explained as Hi-C data becomes available for other plant species with different genomic characteristics.

It is possible that the mechanisms regulating RT in plants differ from those in other systems. The existence of a complete RT profile will be a useful tool to begin filling in these gaps. As discussed earlier, it will be interesting to determine regions that differ in their RT between treatments and compare these regions to genomic features. Elucidation of the mechanisms regulating RT, gene expression and epigenetics in plants is important. Plants are constantly subject to environmental perturbations that they cannot escape. They must be able to adapt to changing conditions while maintaining an overall genomic stability. Global ecology is altered by climate change, population expansion and industrial and agricultural demands are growing. An in depth understanding of the molecular interactions in affecting DNA replication and gene expression in plant systems is necessary to equip scientists with tools to address these issues.
References


Chapter II

Transcriptional and epigenetic shifts during dark adaptation of Arabidopsis cell suspension culture

Introduction

Given the sessile nature of plants, the key to their survival is a genomic plasticity that allows adaptation to changing environmental conditions. This plasticity is achieved by changes in gene expression that are elicited via biochemical pathways in response to various external cues. At the chromatin level, gene expression is regulated by epigenetic modifications that can impact accessibility to transcriptional machinery or take part in recruitment of transcription factors. In particular, modifications to histones represent a reversible epigenetic mechanism that the cell can use to elicit a rapid response when expression or repression of a gene is needed. Tri-methylation of H3K4 (H3K4me3) and H3K27 (H3K27me3) are two important epigenetic modifications that are involved in regulation of gene expression. The H3K4me3 and H3K27me3 marks are regulated by Trithorax group (TrxG) and Polycomb group (PcG) complexes, respectively. The TrxG and PcG complexes were originally studied in a variety of organisms because of their role in regulating developmental genes. However, they are now known to also play a role in the response to environmental cues (Akkers, et al. 2009).

Much ongoing research seeks to establish links between transcription and the mechanisms employed to deposit or remove epigenetic marks on their target genes. A recent model describes how TrxG mediated deposition of H3K4me3 occurs at target genes in Arabidopsis (Song, et al. 2015). The COMPASS-like complex is involved in assembly of the pre-initiation complex (PIC) for transcription as well as deposition of H3K4me3. The authors showed that during transcription of target genes, two sequence-specific transcription factors interact with the COMPASS-like complex, providing a functional link between active transcription and H3K4me3 deposition. On the other
hand functional studies have begun to unveil interactions between chromatin remodelers and PcG proteins that inhibit H3K27me3 deposition (Li, et al. 2015)

Of particular interest is the role of epigenetics in plant adaptation to stressful environmental conditions that threaten the growth and viability of the plant. Many studies have focused on H3K4me3 and H3K27me3 regulated gene expression during stress responses (Kim, et al. 2008; Chinnusamy and Zhu 2009; Kim, et al. 2010; Kim, et al. 2015). Deposition and removal of these marks allows fine-tuning of expression levels in an otherwise unpredictable environment. Under experimental conditions where an imposed treatment is removed, most histone modifications return to their basal level. We do not know how, or if at all, epigenetic marks might be involved in gene regulation during longer periods of adaptation to a new but constant environmental condition. For example, does the epigenetic state remain altered over time in conjunction with minor changes in gene expression in plants that are acclimating to a new environmental condition? Alternatively, does the epigenetic state eventually return to some homeostatic default level once the altered transcriptional state is established?

This is particularly interesting to consider for marks such as H3K4me3 and H3K27me3, which are known to regulate both developmentally and environmentally responsive genes. Genetic pathways regulating environmental responses and development pathways often overlap, which might underlie shared epigenetic regulatory mechanisms (Wasternack and Hause 2002; Nuruzzaman, et al. 2013; Wasternack and Hause 2013). This confounding overlap limits the ability to isolate and only study regulation of environmental effects during adaptation.

Light is one of the most important environmental cues for plants. It provides energy for photosynthesis and is a developmental signal for all stages of the plant life from germination to senescence. Changes in light occur due to natural diurnal and seasonal fluctuations as well as shading by other leaves, plants or physical structures. Plants utilize photoreceptors to intercept light signals and elicit changes in gene expression for
quick and effective respond to changes in light. Many studies suggest that light-responsive genes are regulated by H3K4me3 and H3K27me3 (Guo, et al. 2008; Charron, et al. 2009; Bourbousse, et al. 2015). For example, Guo, et al. (2008) compared mRNA steady-state levels with levels of epigenetic enrichment at several light regulated genes from seedlings grown in constant light and dark conditions for six days. Genes with higher transcripts in the light grown seedlings contained higher levels of H3K4me3 than dark grown seedlings. The light up-regulated genes had higher levels of H3K27me3 in the dark grown seedling. Many plant light based studies comparing H3K4me3 and H327me3 levels with gene expression are performed in whole plants. Thus the epigenetic and gene expression changes are related directly to the environmental response as well as development responses elicited by the light.

We do not know if changes in enrichment of H3K4me3 and H3K27me3 in differentially expressed genes are maintained or if they return to basal levels over long periods of time. The idea that there could be a return to basal levels is supported by the fact that histone modifications are known to be reversible, and the possibility that there may be an energy cost to continually maintain a non-basal state, particularly in a cell line that is not subject to developmental changes.

This chapter describes a set of experiments designed to answer these questions. We compared transcriptional and epigenetic changes between light and dark grown Arabidopsis suspension cells. Our light grown cells are green and grown under constant light. A dark-adapted line was derived from the light grown cells by maintaining them in constant dark conditions for two weeks. This resulted in cell lines that were genetically identical and differed only in their growth conditions. This allowed us to focus on gene expression and epigenetic changes resulting from differences in environmental treatment, and largely avoid developmental effects that can complicate data generated from whole plants, which contain a mixture of cell types, each with a different transcriptional program, which can vary considerably in
abundance relative to one another during a major developmental process such as leaf development. Changes in H3K4me3 and H3K27me3 abundance and localization were analyzed by ChIP-Seq and compared to RNA-Seq data to determine if these epigenetic modifications are correlated with differential gene expression during a period of adaptation to dark conditions.

**Results and Discussion**

*Phenotypic and gene expression differences*

To generate a dark adapted Arabidopsis cell culture, light grown suspension cells were placed in continuous dark conditions and sub-cultured into fresh medium every seven days. After two weeks the cell suspension appeared pale yellow, suggesting a change in expression level of genes involved in photosynthesis and chlorophyll metabolism (Figure 1). Thus, we considered cells grown under constant dark for two weeks to be dark-adapted. Replicating cells were collected 16 hours after sub-culture into fresh medium. The 16 hour time point was chosen based upon a preliminary study which found that BrdU incorporation into DNA is maximized at this time (Lee et al 2010). Total RNA treated with DNase and depleted of ribosomal RNA was used to generate RNA-Seq libraries for three bioreps of light and dark adapted cells. The RNA-Seq samples were sequenced on an Illumina instrument. RNA-Seq reads were aligned to the Tair 10 genome with TopHat2. An average of 3,987,744 and 2,328,545 mappable paired-end reads were obtained from the light and dark samples, respectively (Table 1).

Differential gene expression analysis of the RNA-Seq data was performed using EdgeR. A false discovery rate of <0.05 and a fold-change of ≥2 were used to classify genes that are significantly different between the light and dark treatments. The analysis revealed that 2.3% of genes were differentially expressed. A total of 239 genes were up-regulated and 379 genes were down-regulated in the dark relative to the light. MapMan category analysis determined that secondary metabolism was the most over-represented gene family (P-value = 6.65e-07) in the up-regulated gene set (Table 2).
This result most likely reflects the shift in metabolic activity required for energy metabolism in the dark. Accordingly, photosynthesis and tetrapyrrole gene families were over-represented (P-value = 1.47e-67 and 5.89e-08, respectively) in the down-regulated gene set, which is most likely due to a reduction in photosynthetic metabolism and chlorophyll production. This is supported by the phenotypic change from green to white cells that occurs in the absence of light.

**Genome-wide enrichment of H3K4me3 and H3K27me3**

To study enrichment of H3K4me3 and H3K27me3, genome-wide profiles for each mark were generated using ChIP-Seq. Replicating cells were collected 16 hours after sub-culture into fresh medium. This time point was chosen based upon preliminary work showing that maximal nucleotide analog incorporation is achieved after 16 hours (Lee, et al. 2010). Chromatin immunoprecipitation was performed using antibodies recognizing H3K4me3 and H3K27me3. Immunoprecipitated DNA and a genomic DNA reference were sequenced. An average of 32,687,473 and 37,572,529 reads were mapped to the Tair 10 genome using the Burrows-Wheeler Aligner (BWA). The normalized fold enrichment (FE) for each mark was calculated as the ratio of mapped reads from the ChIP sample to mapped reads from the genomic DNA reference in the same genomic interval. H3K4me3 and H3K27me3 are typically genic marks known to be involved in active and repressive gene expression, respectively. We first asked if their patterns of enrichment across genes in the suspension cell lines are similar to those seen in other experimental systems and determined if the genome-wide patterns of enrichment are similar between the light and dark treatments. To address these questions, whole genome average gene analysis of enrichment was performed. First the average FE was calculated across all genes. To normalize for length, averages were calculated at 50 equal bins across the gene body. The average FE was also calculated for regions 1 kb upstream from the transcription start site and 1 kb downstream from the beginning of the 3’ UTR. The average FE in the 1 kb upstream and downstream regions
were calculated in 50 equal bins. In accordance with other studies, H3K4me3 was enriched at the 5' end of genes while H3K27me3 was more broadly distributed across genes (Figure 2). It should be noted that the mean H3K27me3 FE across genes is < 1, indicating a depletion of this mark on genes compared to the genomic background level. The reason for this depletion is that only 22% of genes were enriched for H3K27me3 (Table 3). The remaining 78% of genes did not contribute to the H3K27me3 signal, thereby reducing the average enrichment.

*Integration of differential gene expression and ChIP-Seq data*

Shifts in epigenetic patterns often coincide with changes in gene expression in different treatments. We examined the level of difference in the H3K4me3 and H3K27me3 profiles between light and dark grown cells using the MACS2 callpeak module to identify genomic regions where there was a loss or gain in epigenetic enrichment after the two-week dark adaptation. In the light, 22,132 peaks of H3K4me3 enrichment were called (Table 4). The dark samples contained 20,828 peaks. Most of H3K4me3 peaks (91.1%, 97.5%) were shared between light and dark. There was a 7.9% unique loss and a 2.5% unique gain of H3K4me3 peaks in the light to dark transition. For H3K27me3, 7,127 and 7,058 peaks were called in light and dark, respectively. Similar to the H3K4me3 results, most of the H3K27me3 peaks (92.2%, 94.1%) were shared between light and dark. There was a 7.8% unique loss and a 5.9% unique gain of H3K27me3 regions of enrichment in the light to dark transition. Overall, the dark samples had slightly fewer unique H3K4me3 and H3K27me3 regions than the light samples but the H3K4me3 and H3K27me3 profiles were highly similar.

Figure 3a displays a histogram of the distribution of overlap between differentially expressed genes and H3K4me3 peaks, including unique (green or purple bars) and shared (grey bars) peaks. The histograms also display the count of differentially expressed genes that did not contain a mark (white bars). Of the 379 genes that were significantly down-regulated in dark, 91 (24%) genes were marked by H3K4me3 peaks
only enriched in light samples, 249 (66%) genes were in regions that were H3K4me3 enriched in both light and dark samples, and 39 (10%) genes were not enriched for H3K4me3 in either light or dark. None of the genes down-regulated in dark contained a H3K4me3 peak that was unique to the dark profile.

Of the 239 genes that were significantly up-regulated in dark, 18 (0.1%) genes were marked by H3K4me3 peaks only enriched in the dark samples, 4 (< 0.1%) genes were marked by peaks only in the light samples, and 167 (69%) genes were marked by H3K4me3 in light and dark samples. Fifty of the up-regulated genes did not contain an H3K4me3 mark in either the light or dark samples.

The overlap of H3K27me3 marks with differentially expressed genes displayed a different distribution than H3K4me3 (Figure 3b). Most of the differentially expressed genes did not contain an H3K27me3 mark (white bars). Sixteen genes that were up-regulated in dark were uniquely marked by H3K27me3 in the light while only one gene was marked in the dark. A totally of 52 up-regulated and 45 down-regulated genes were marked by H3K27me3 in both the light and dark samples (grey bars).

Initially we did not expect to see a high proportion of differentially expressed genes maintaining similar number of H3K4me3 and H3K27me3 peaks between the light and dark samples. However, closer inspection of the data provided insight into these unexpected results. Figure 4 displays two IGV browser shots that contain RNA-Seq, H3K4me3 and H3K27me3 data for the light and dark treatments. In Figure 4a, the blue (dotted lines) box contains sequencing data for the Arabidopsis gene AT3G56940, which is down-regulated in the dark samples (log2FC = -4.01 and P-value = 7.42e-23). The first two tracks support the differential gene expression results as the RNA-Seq read pile-up is greater for light than dark. The third and fourth tracks show that H3K4me3 is enriched in both light and dark; accordingly these regions were identified by peak calling as indicated by the black bar beneath each peak. Subsequently both regions (sets of peaks) overlapped the differentially expressed gene in our analysis.
Although the light sample maintains a higher amount of H3K4me3 at this gene, the peak called in the dark sample appears to be real.

Another type of example is illustrated in Figure 4b. The gene in this example was up-regulated in the dark (log₂FC = 7.67 and P-value = 4.65e-25). A comparison of the H3K4me3 tracks reveals that there is a higher level of the histone mark in the dark, and in agreement with the visual data, a peak was called in the dark but not the light. However, visual comparison of H3K27me3 tracks shows that the light sample has a higher level of the mark than the dark sample, which is what we expect given that the gene is up-regulated in dark. However, the peak caller did not identify this region as enriched for H3K27me3 in either light or dark samples. It is not uncommon for peak callers to miss peaks that are recognized by the human eye (Strino and Lappe 2016). Rye, et al. (2011) speculate that algorithms employed to call peaks cannot completely utilize the visual cue of peak shape. Misinterpretation of data caused by anomalies such as missed peaks can be amplified when comparing across data types.

A second approach was used to examine the changes occurring during dark adaptation. In brief, the genome-wide across gene analysis of epigenetic enrichment described earlier was adapted to analyze enrichment patterns in only those genes that were differentially expressed. The average H3K4me3 and H3K27me3 FE was calculated for 50 equal bins across the genes, as well as the 1 kb upstream and downstream regions. The average epigenetic FE was plotted to study enrichment pattern differences between up- and down-regulated genes for the light and dark samples. Genes that were up-regulated in dark had slightly higher 5’ H3K4me3 levels in dark samples than in light (Figure 5a). Genes that were down-regulated maintained higher H3K4me3 levels in the light than those genes in the dark (Figure 5b). Genes that were up-regulated in dark maintained higher levels of H3K27me3 in light than in dark whereas the opposite pattern was observed for the down-regulated genes (Figure 5c,d).
Comparison of Figure 5a with Figure 2a revealed that the up-regulated genes maintained lower levels of H3K4me3s in both the light and dark treatments than the levels across all genes. For down-regulated genes, levels were lower than the signal for all genes only in the dark samples. The up-regulated genes maintained a noticeably greater level of H3K4me3 in light samples relative to the dark samples, which was comparable to the genome-wide level measure shown in Figure 2a. Thus, with the exception of down-regulated genes in the light samples, the differentially expressed genes maintained H3K4me3 levels below the average calculated for all genes.

Guo, et al. (2008) presented a conclusion regarding the comparison of epigenetic levels with gene expression that could relate to our findings. They proposed that the degree of change in epigenetic levels is less than the degree of change in expression at genes that are regulated by a particular mark. This notion could also relate to our observation that epigenetic changes that are too small to detect by differential peak calling might influence gene expression changes that are quite large and detectable in a differential expression analysis.

**Conclusions**

This research asked H3K4me3 and H3K27me3, two epigenetic marks involved in regulating gene expression, are maintained on responsive genes after an extended period of adaptation to dark conditions. This study was unique for two reasons. 1) It avoided strong developmental differences that impact many genetic pathways. 2) We examined these relationships after an extended time period of adaptation versus after a shorter exposure time.

We found that the several differentially expressed genes also exhibited differential enrichment of H3K4me3 and H3K27me3. We initially used peak calling as means to relate epigenetic changes with genes expression. However, we demonstrated that it is necessary to explore both qualitative and quantitative differences.
Materials & Methods

Cell growth conditions

The *A. thaliana* suspension cell line Col-0 (ecotype Columbia) was grown in 250 mL flasks containing 50 mL culture medium [3.2 g/L of Gamborg’s B5 basal medium with minimal organics, 3 mM MES, 3% sucrose, 1.1 mg/L 2,4-D, pH 5.8]. Cell flasks were maintained under constant light at 23°C on a rotary shaker at 160 rpm. Cells were sub-cultured every 7 days by aliquoting 6 mL of the cell suspension into 50 mL fresh culture medium. To maximize the number of cells in the logarithmic phase of growth at the time of assay, 25 mL of 7-day old cells were transferred to 25 mL of fresh medium and allowed to grow for 16 hours prior to harvesting.

For the dark treatment, all conditions were the same as described above with the exception that cells were grown in constant dark for two weeks. The dark treatment was applied by growing cells in flasks that were completely covered with aluminum foil. At day 7, a 6 mL aliquot of the dark grown cells was rapidly transferred under low light conditions to fresh medium and maintained in the dark for a second week of adaptation. After seven days, 25 mL of dark-adapted cells culture was placed into 25 mL of fresh medium and the cells were grown in dark for 16 hours.

RNA-Seq sample preparation

Cells were collected at 16 hours after sub-culturing into fresh medium as described. One flask of cells comprised a single bioreplicate and three bioreplicates were collected. Cells were washed three times with cold 1x phosphate-buffered saline (PBS), snap-frozen with liquid nitrogen and stored at -80°C. Frozen cells were ground in a mortar and pestle with liquid nitrogen. Total RNA was extracted from 10 μg of frozen ground cell material using PureLink Plant RNA Reagent (Ambion). Residual DNA was removed using Turbo DNA-free kit (Ambion), yielding 7.6 μg of RNA (~76% recovery of RNA). Ribosomal RNA (rRNA) was depleted from 4 μg of DNA-free RNA using the Ribo-Zero
Magnetic Kit for Plant Leaf material (EpiCentre). The efficiency of rRNA depletion was determined by quantitative real-time polymerase chain reaction (qRT-PCR) using primers for the Arabidopsis 5S, 18S, 23S and 25S rRNAs. qRT-PCR was performed with qScript One-Step SYBR Green qRT-PCR kit, Low ROX (Quanta BioSciences). RNA-seq libraries were prepared with the ScriptSeq™ v2 RNA_seq Library Preparation Kit (EpiCentre). Each sample was uniquely barcoded with ScriptSeq™ Index PCR Primers (EpiCentre) and samples were pooled for sequencing on one lane of an Illumina HiSeq 2000 instrument.

**Epigenetic ChIP-Seq assays**

Cells were collected at 16 hours after sub-culturing into fresh medium. A total of four bioreplicates were collected for each epigenetic ChIP-Seq assay. At the 16-hour time point, cells were crosslinked with 1% paraformaldehyde for 10 minutes at room temperature and washed three times with cold 1x PBS. The samples were snap-frozen with liquid nitrogen. The frozen cell pellet was ground to a fine powder in a mortar and pestle and stored at -80°C.

Chromatin immunoprecipitation (ChIP) was performed on the frozen ground cell material according to Gendrel, et al. (2005) with modifications. Chromatin was sonicated using the S2 Covaris with the parameters (duty = 10%, intensity = 5, bursts/sec = 200) optimized to obtain a final DNA fragment size of 200-300 bp. The sonicated chromatin was precipitated overnight at 4°C with either an H3K4me3 (Millipore, Billerica, Massachusetts, USA) or H3K27me3 (Millipore) antibody. The immunoprecipitates were washed, the crosslinks were reversed, and the samples were purified using the Qiagen PCR Purification kit. The NEXTflex Illumina Chip-Seq Library Prep Kit with the ultra-low input protocol was used to generate sequencing libraries. The libraries were sequenced on an Illumina HiSeq 2000 instrument.
**Bioinformatics analysis**

FastQC /0.10.1 was used to assess the quality of the raw sequencing data. RNA-seq and ChIP-seq reads were trimmed using Trim Galore! to remove Illumina Truseq adapter index sequences used in multiplexing of samples (Krueger 2015). A read pair was discarded from further analysis if it contained a read with less than 40 bp after trimming.

**Differential expression analysis**

Processed RNA-Seq reads were mapped to the Arabidopsis TAIR 10.0 genome using TopHat2 v2.0.13. Duplicate and improperly paired reads were removed using Samtools v0.1.19. Fragments Per Kilobase of exon per Million fragments mapped (FPKM) values were generated using Cuffnorm v2.2.1. The HT-Seq python program was used to calculate read counts for genes and EdgeR was used to estimate differential gene expression. The criteria to designate differential expression were a FDR < 0.05 and a fold-change of > 2.0.

**Epigenetic ChIP-Seq**

ChIP-Seq reads were mapped to the TAIR 10.0 version of the Arabidopsis thaliana genome using the Burrows-Wheeler Alignment package with the bwa-mem algorithm (bwa/0.7.12) (Li and Durbin 2009). The deepTools package was used to assess correlation between biological replicates (Li, et al. 2009; Ramirez, et al. 2014). After correlation analysis was performed, ChIP bam files of each biological replicate were merged with samtools.

MACS2 v2.1.0 was used to analyze ChIP-Seq data (Liu 2014). To call peaks of enrichment and to generate normalized signal per million reads (SPMR) bedgraph files, the macs2 callpeak module was used with default parameters and the following additional flags: --SPMR --broad --broad-cutoff 0.01. Fold enrichment of the ChIP sample
to G1 input control was calculated using the normalized SPMR bedgraphs in the macs2 bdgcmp module with default setting and the following flag: \textit{-m FE}. 
References


Figure 1. Phenotypic comparison of light and dark grown cells. Cells were considered to be dark-adapted after two weeks in constant dark conditions.
Figure 2. Mean H3K4me3 signal (a) and H3K27me3 signal (b) across genes +/- 1 kb in light and dark samples. The grey dotted line indicates IP to control ratio signal = 1 to delineate enrichment above and below background.
Figure 3. Histograms displaying overlap between differentially expressed genes and H3K4me3 or H3K27me3 peaks. Histogram bars are color coded as follows: green = count of differentially expressed genes that contained an epigenetic peak only the light samples, purple = count of differentially expressed genes that contained an epigenetic peak only the dark samples, grey = count of differentially expressed genes that contained an epigenetic peak in light and dark, white = count of differentially expressed genes that did not contain an epigenetic peak in light or dark.
Figure 4. IGV browser shots illustrating relative RNA Seq, H3K4me3 and H3K27me3 fold-enrichment tracks for light and dark. The relevant down regulated (a) and up-regulated (c) gene regions are surrounded by a broken blue box. A solid black line beneath a peak indicates a region of epigenetic mark enrichment as determined by the MACS2 peak caller.
Figure 5. Mean H3K4me3 signal (a,b) and H3K27me3 signal (c,d) across genes -/+ 1kb that are up-regulated (a,c) or down-regulated (b,d) in the dark.
Table 1. Mappable read counts from Illumina paired-end data.

<table>
<thead>
<tr>
<th></th>
<th>Biorep</th>
<th>Light</th>
<th>Dark</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RNA Seq</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2163613</td>
<td>2446007</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3388255</td>
<td>2724444</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6411363</td>
<td>1815183</td>
<td></td>
</tr>
<tr>
<td><strong>H3K4me3 ChIP Seq</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>42647765</td>
<td>26315367</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>37944687</td>
<td>29737139</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>37315965</td>
<td>32629471</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>40730119</td>
<td>38271112</td>
<td></td>
</tr>
<tr>
<td>Input</td>
<td>29224111</td>
<td>36474278</td>
<td></td>
</tr>
<tr>
<td><strong>H3K27me3 ChIP Seq</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>47585034</td>
<td>43642450</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>44386074</td>
<td>28444187</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>33817532</td>
<td>39049038</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>52698114</td>
<td>34743808</td>
<td></td>
</tr>
<tr>
<td>Input</td>
<td>43856509</td>
<td>35444977</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Enriched MapMan category terms with corrective P-value ≤ 0.01 of genes up-regulated and down-regulated in dark

<table>
<thead>
<tr>
<th>MapMan category term</th>
<th>Representation</th>
<th>Corrected P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dark: up-regulated</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary metabolism</td>
<td>over</td>
<td>6.65e-07</td>
</tr>
<tr>
<td>Misc.</td>
<td>over</td>
<td>2.32e-06</td>
</tr>
<tr>
<td>RNA</td>
<td>over</td>
<td>6.79e-03</td>
</tr>
<tr>
<td>Minor CHO metabolism</td>
<td>over</td>
<td>0.01</td>
</tr>
<tr>
<td>Signaling</td>
<td>over</td>
<td>0.03</td>
</tr>
<tr>
<td>Not assigned</td>
<td>under</td>
<td>5.51e-04</td>
</tr>
<tr>
<td>Protein</td>
<td>under</td>
<td>8.47e-04</td>
</tr>
<tr>
<td><strong>Dark: down-regulated</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Photosynthesis</td>
<td>over</td>
<td>1.47e-07</td>
</tr>
<tr>
<td>Tetapyrrole synthesis</td>
<td>over</td>
<td>5.89e-08</td>
</tr>
<tr>
<td>Secondary metabolism</td>
<td>over</td>
<td>4.04e-04</td>
</tr>
<tr>
<td>Misc.</td>
<td>over</td>
<td>3.53e-03</td>
</tr>
<tr>
<td>Transport</td>
<td>over</td>
<td>6.25e-03</td>
</tr>
<tr>
<td>Lipid metabolism</td>
<td>over</td>
<td>0.01</td>
</tr>
<tr>
<td>Co-factor/vitamin metabolism</td>
<td>over</td>
<td>0.01</td>
</tr>
<tr>
<td>Stress</td>
<td>over</td>
<td>0.01</td>
</tr>
<tr>
<td>Not assigned</td>
<td>under</td>
<td>3.28e-09</td>
</tr>
<tr>
<td>Protein</td>
<td>under</td>
<td>7.71e-04</td>
</tr>
</tbody>
</table>
Table 3. Overlap of MACS2 peaks and genes.

<table>
<thead>
<tr>
<th>Mark</th>
<th>Condition</th>
<th>Total peaks</th>
<th>Peaks overlapped by a gene</th>
<th>Genes overlapped by a peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K4me3</td>
<td>Light</td>
<td>22132</td>
<td>18869 (84%)</td>
<td>18843 (85%)</td>
</tr>
<tr>
<td></td>
<td>Dark</td>
<td>20832</td>
<td>17905 (84%)</td>
<td>17905 (86%)</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>Light</td>
<td>7127</td>
<td>4591 (64%)</td>
<td>6076 (22%)</td>
</tr>
<tr>
<td></td>
<td>Dark</td>
<td>7058</td>
<td>4519 (64%)</td>
<td>5980 (22%)</td>
</tr>
</tbody>
</table>
Table 4. Comparison of MACS2 peak calling between light and dark.

<table>
<thead>
<tr>
<th>Condition</th>
<th>H3K4me3 peaks</th>
<th>H3K27me3 peaks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Unique (%)</td>
</tr>
<tr>
<td>Light</td>
<td>22132</td>
<td>1764 (7.9)</td>
</tr>
<tr>
<td></td>
<td>7127</td>
<td>5617 (7.8)</td>
</tr>
<tr>
<td>Dark</td>
<td>20832</td>
<td>5142 (2.5)</td>
</tr>
<tr>
<td></td>
<td>7058</td>
<td>4191 (5.9)</td>
</tr>
</tbody>
</table>
Chapter III

Epigenetic state and transcription are related to replication timing in Arabidopisis

Introduction

DNA replication occurs in the synthesis phase (S-phase) of the cell cycle during which the entire genome is duplicated before mitotic cell division. Replication initiates at genomic locations termed “origins” and bidirectional replication forks proceed away from an origin until they reach a termination zone. The span of DNA that is replicated from a single origin to the termination zones forms a replicon. The ability to study replicons was made possible by the development of radioactively labeled or otherwise modified nucleosides that can be converted into nucleotides by the cell and incorporated into a growing DNA strand during synthesis (Taylor 1957). These synthetic nucleosides enable visualization or isolation of newly replicated DNA. In the 1990s, cytological visualization of 5-bromodeoxyuridine (BrdU) in corporation revealed the presence of replicon clusters or replication foci, which replicated at the same time and occupied distinct spaces in the nucleus during the cell cycle (Sparvoli, et al. 1994; Jackson and Pombo 1998). Replication foci that were labeled early in S-phase maintained early replication in the daughter cells through several cell cycle phases suggesting that this shared timing was a stable, heritable feature of replication.

It is now well established that eukaryotic DNA replication occurs in a well-defined temporal order, referred as the “replication timing” (RT) program. In all eukaryotes studied, RT patterns are conserved in the same cell type of a species. For example, human embryonic stem cells (hEScs) from several different cell lines have similar profiles (Ryba, et al. 2010). Furthermore, RT patterns are conserved across species for certain cell types. For example, RT in human cells is similar to RT patterns in murine cells at a corresponding developmental stage (Ryba, et al. 2010). These measures of
similarity are based on comparison of RT domain size, timing of orthologous genes and patterns observed at chromosomal regions of synteny.

During cell lineage determination large changes in the RT program occur (Hiratani, et al. 2008; Hiratani, et al. 2010). Regions of the genome that change RT often display a corresponding change in nuclear localization suggesting that the spatial arrangement of chromatin in the nucleus is important for regulation of replication. It also suggests that establishment of a RT program is governed by principles of genomic organization (that are related to or impact the epigenetic state or transcription). A large body of research comparing RT data to epigenetic and transcriptional data supports this notion (Rivera-Mulia and Gilbert 2016; Sequeira-Mendes and Gutierrez 2016).

Associations between early replicating regions and increased transcriptional activity in Drosophila, murine, human and plant systems have been reported (Schubeler, et al. 2002; MacAlpine, et al. 2004; Lee, et al. 2010). Conversely, late replicating regions have been associated with decreased gene expression (Rivera-Mulia and Gilbert 2016). Early replication is associated with epigenetic marks for open chromatin such as H4K16ac and depleted in heterochromatic marks such as H3K9me2 (Schwaiger, et al. 2009; Lee, et al. 2010). The H3K27me3 epigenetic mark, which is involved in transcriptional repression, is enriched in late replicating regions. Together, these observations provide a link between RT, gene expression and epigenetic regulation (Thurman, et al. 2007; Bell, et al. 2010). However, the simplistic association of early and late RT with high and low levels of transcription, respectively, is violated in many cases, thereby precluding a direct cause and effect relationship between the processes.

The development of chromosome conformation capture (3C) techniques has contributed greatly to our understanding of the link between 3-dimensional structure and function of chromatin processes. (Dekker, et al. 2013). The 3C techniques identify long-range interactions between regions of the genome that are often of biological relevance (Dekker, et al. 2002; Lieberman-Aiden, et al. 2009). The Hi-C technique,
which uses 3C technology with high throughput sequencing, revealed the presence of strong interactions between loci in large, local regions of chromatin named “topologically associated domains” (TADs) in mammals and Drosophila (Dixon, et al. 2012; Sexton, et al. 2012). TADs are bordered by insulators, suggesting a potential association of the chromatin structure with transcription. Furthermore, comparison of mammalian interaction data with RT profiles revealed the existence of a correlation between TAD boundaries and RT domains (Pope, et al. 2014). These findings provided strong evidence for the hypothesis that RT is regulated by sub-nuclear chromatin dynamics and also highlighted a potential functional link between RT and transcription.

The 3C patterns observed in Arabidopsis differ from those in other species, indicating that Arabidopsis might use unique mechanisms to regulate chromatin interactions (Grob, et al. 2013; Feng, et al. 2014; Wang, et al. 2015). However, there were features of the Arabidopsis interaction that suggest mechanisms of regulation are shared. A Hi-C study in Arabidopsis identified several strong interactions referred to as “positive strips” (Wang, et al. 2015). Positive strips were enriched for the repressive mark, H3K27me3, and genes that fell into the positive strips were transcribed at lower levels than genes outside of these regions. Moreover, regions adjacent to the positive strips displayed weaker interactions and, thus, designated as “insulator-like”. These insulator-like regions are enriched for active epigenetic marks and highly expressed genes.

Despite expansive research on RT in yeast, mammals and Drosophila, research in plants has been in part hindered by polyploid events that result in gene duplications (Lee, et al. 2010). Functional divergence of duplicated genes over times makes it difficult to study their role in some cellular processes (Roulin, et al. 2013). Nonetheless, Arabidopsis has emerged as an excellent plant to study RT due to its small, well-annotated genome that contains significantly fewer repeat regions than other plant species. Lee et al (2010) used pulse labeling with BrdU and one-dimensional (1D) fluorescence activated cells sorting (FACS) based on DNA content (DAPI) to sort early
S/G1, mid S and late S/G2 populations. Early, mid and late S replicating DNA from the sorted populations were immunoprecipitated with an antibody against BrdU. RT profiles of chromosome IV were generated using tiling microarray analysis of DNA from S-phase and a G1 input control. They also mapped H3K56 acetylation on chromosome 4 and showed that it is associated with early RT in Arabidopsis (Lee, et al. 2010).

To investigate the relationship between replication timing, epigenetics and transcription we have generated a genome-wide RT profile and compared it with epigenetic and transcriptional data from the same Arabidopsis cell line grown under the same conditions. In our experiments, Arabidopsis cells were pulse labeled with the nucleoside 5’-Ethynl-2’-deoxyuridine (EdU) and labeled DNA was conjugated to Alexa-fluor (AF488). Two-dimensional (2D) FACS based on EdU incorporation/AF488 fluorescence and DNA content was used to obtain enriched population of early, mid and late S-phase fractions and a G1 fraction from a mixed cell population. EdU-labeled DNA immunoprecipitated from the S-phase populations was sequenced and mapped. Total DNA from the G1 population was sequenced and used as an input control. A recently developed RT segmentation algorithm used the sequencing maps to generate a whole genome RT profile for Arabidopsis (Zynda, et al. 2016). Epigenetic enrichment profiles were generated for H3K4me3, a euchromatic mark for active transcription, H3K27me3, a heterochromatic mark associated with repressed transcription, and H3K56ac, a euchromatic mark that has been associated with early replication timing in Arabidopsis.

**Results and Discussion**

*Replication timing basics: Distribution of classes across the genome, segment sizes*

To generate genome wide replication timing profiles of Arabidopsis, cells at 16 hours post 7-day split were labeled with EdU for 30 minutes, fixed and flash frozen. The 16 hour time point was chosen based on preliminary experiments that determined this time to be optimal for DNA labeling. Nuclei were isolated, AF488 was conjugated to EdU and total DNA was stained with BrdU. Nuclei were separated into early, mid and
late S-phase and G1-phase by 2D FACS based on EdU/AF488 and BrdU. This approach differed from the previous tiling array analysis of Arabidopsis chromosome IV which used 1D FACS based on DNA content and BrdU immunoprecipitation of labeled replicating sequences. Figure 1 is an Arabidopsis FACS profile. The y-axis of the plot indicates the relative level of EdU incorporation as determined by AF488 fluorescence and the x-axis indicates total DNA content. The gating strategy used to isolate the cell phase fractions for RT analysis is indicated by the red boxes. EdU-labeled DNA from each S-phase fraction was immunoprecipitated using an antibody to AF488, sequenced and mapped to the Arabidopsis Tair10 genome. The G1 population of nuclei served as a whole genome control where each locus is present at the 2C level. Table 1 shows the mappable read counts obtained for S-phase fractions and G1-phase. The RT profiles were generated by normalizing mapped read count density from each S-phase fraction at each locus to the G1 control read count for that locus. Figure 2 illustrates the early, mid and late S-phase smoothed profiles for each of the five Arabidopsis chromosomes. The genome wide view highlights global differences between the profiles. The top track is a map of early replicating sequences that clearly shows that early replication is enriched towards the ends of chromosomes and depleted in centromeric regions. The middle S-phase sequences in the second track are evenly distributed across the genome while late replicating sequences are enriched in the centromere region.

Most previous studies have assigned replication times using a ratio of early to late signal for the locus in question. The use of EdU labeling with FACS provided us with the unique opportunity to assess signals directly from three fractions of S-phase. A novel algorithm (Zynda, et al 2016) was used to assess enrichment of sequences from each population, and every 1-kb region of the genome was segmented into replication timing classes of either "Early" (E), "Early-Mid" (EM), "Mid" (M), "Mid-Late" (ML), "Late" (L), Early-Late” (EL) or “Early-Mid-Late” (EML). Regions of the genome that were indeterminate (i.e. low sequencing coverage) were assigned as "NA". Segmentation was performed by dividing the genome into bins and the early, mid and late signals were
compared within each bin to assign a RT class. For some loci replication can occur in more than a single fraction of S-phase which results in two predominate RT signals For example, the signals from the early and mid populations could be similar and much greater than the late signal in a given bin. Although the bin in this example would not be considered “late”, it would not be possible to assign it as “early” or “mid”. Instead, it would be considered half way in between early and mid or “early-mid”. Similarly, a bin that was not represented in the early sample and had similar sequencing signals in the mid and late samples was classified as “mid-late”. Assignments of early-late and early-mid-late were also possible.

The distribution of replication timing is similar across each of the five chromosomes (Figure 4a). Similar to the pattern seen Figure 3, the centromeric region of each chromosome is L (red) with a transition to earlier RT moving away from the centromere towards the distal arms. This pattern of RT coincides with earlier reports of spatially distinct interactions of heterochromatic pericentromere and centromere regions into “chromocenters” and loops of euchromatin that do not interact (Fransz, et al. 2002). One exception we observed was the presence of a L region (~500 kb) at the distal short arm of chromosome 2 (Figure 5) which contains the heterochromatic 45S rDNA nucleolar organizing region (NOR) (Fransz, et al. 2002). Interestingly, the NOR located at the short arm of chromosome II was shown to interact with chromocenters, suggesting that RT in Arabidopsis is related to chromatin structure and sub-nuclear positioning.

There was a substantial amount of variability in the distribution of timing classes and in the size of the replicating segments across each of the timing classes (Figure 4b,c). The pie chart shows that most of the Arabidopsis genome replicates in E and EM. Furthermore, E and EM timing segments were larger than M, ML and L segments (Figure 4b). This could reflect that the Arabidopsis genome is highly euchromatic and gene dense. Mid timing segments were small and represented 10% of the genome. Mid-
late segments were larger than M and L and were also relatively abundant (~15%). Late segments were smaller in size and represented only 4% of the genome. The EL and EML timing classes were infrequent and evenly distributed across the genome (Figure 4b). Based upon the negligible amount of EL and EML in the genome, these timing classes were excluded from downstream analyses.

*Replication timing distributed across epigenetic state*

The epigenome impacts the chromatin environment using various mechanisms to affect the accessibility of DNA for transcription and replication. It has been hypothesized that epigenetic regulation plays a role in initiation of replication and determines replication time. Thus, one of our goals was to study the relationship between epigenetics and RT in Arabidopsis. To this end, we examined a small combination of epigenetic marks that are known to be relevant to transcription and/or RT in Arabidopsis by comparing their enrichment across RT classes. The H3K4me3 and H3K27me3 marks were selected for regulatory roles in active and repressive transcription, respectively. The euchromatic mark, H3K56ac, was chosen because it is associated with early RT and is associated with nucleosome dynamics in Arabidopsis (Lee, et al. 2010; Watanabe, et al. 2013).

Epigenetic ChIP-Seq was performed to generate genome-wide maps for each mark. We first assessed the enrichment of the histone modifications across the RT classes. Observation of the general trends of epigenetic enrichment revealed several interesting aspects of the RT data (Figure 6). In general, H3K4me3 enrichment was highest for E and EM replicating regions of the genome. Mid and ML replicating sequences contained lower levels that were similar to each other while L replicating sequences had the lowest levels of H3K4me3. A similar pattern was observed across the RT classes for H3K56ac. E and EM were enriched above the genome control (i.e. ratio > 1.0) for H3K4me3 and H3K56ac and were depleted in M, ML and L. The opposite pattern was observed for H3K27me3 with respect to RT class. These results are similar to findings
in other organisms and suggest a link between early RT, active transcription and euchromatin in Arabidopsis (Ryba, et al. 2010; Comoglio and Paro 2014).

Interestingly, H3K27me3 was most highly enriched in M and levels decreased as replication timing became later. Early and early-mid regions maintained the lowest levels of H3K27me3 enrichment. In mammalian systems, H3K27me3 is highest in heterochromatic, late replicating regions. However, most studies measure RT by comparing a ratio of early to late, which precludes distinguishing placement of epigenetic marks during multiple times during S-phase. The segmentation approach used in our study allowed us to analyze finer slices replication, revealing that H3K27me3 is highest in M. The fact that H3K27me3 levels are lower in ML and L than in M, supports its role as a facultative mark for heterochromatin and suggests that H3K27me3 is not actively driving changes in RT (Makarevitch, et al. 2013). The H3K9me2 modification is a mark for constitutive heterochromatin (Roudier, et al. 2011). It would be interesting to determine if this mark is high in later replicating sequences.

Empirical data and mathematical modeling predict that the combinatorial activity of epigenetic marks better characterizes chromatin state than the presence of a single epigenetic mark (Comoglio and Paro 2014). These conclusions are based on analyses of a large combination of marks for which we do not have data for comparison to our RT profiles. Nonetheless, we were interested to determine if any notable patterns emerged when our RT data was analyzed in the context of any combination of H3K4me3, H3K27me3 and H3K56ac. To begin this analysis, the genome was binned into 1kb windows. Each window was classified as either “enriched” (IP:control ratio >1) or “depleted” (IP:control ratio <1) for each of the three epigenetic marks. The analysis of three marks resulted in seven distinct combinations of epigenetic marks, termed epigenetic signatures (Table 2). For example, a window enriched for H3K4me3 but depleted in H3K27me3 and H3K56ac would have an epigenetic signature of only
“H3K4me3”. A window enriched for H3K4me3 and H3K56ac but depleted in H3K27me3 would have an epigenetic signature of “H3K4me3 + H3K56ac”.

The number of bins belonging to each RT class was counted for each epigenetic signature. A chi-square test was performed to determine if the RT classes were enriched for the different epigenetic signatures (Table 3). The distribution of RT classes across the epigenetic signature groups is shown in Figure 7. The epigenetic groups containing H3K56ac or H3K56 + H3K4me3 were enriched for E. However, groups with H3K4me3 alone were not significantly enriched for E. The groups enriched for H3K4 alone or in combination with H3K56ac were significantly enriched for EM. This suggests that H3K56ac marks regions of the genome that replicate early in S-phase. Histone acetylation facilitates de-compaction of chromatin by loosening DNA interaction with the nucleosome. The association between early replication and H3K56ac could reflect nucleosome displacement or incorporation of the histone variant H2A.Z. The H2A.Z variant is enriched near Arabidopsis replication origins (Costas, et al. 2011). Furthermore, H3K56ac regulates exchange of H2A and H2A.Z via interactions with a chromatin modifier (Watanabe, et al. 2013). Although H3K4me3 is associated with earlier RT in general, its presence alone is not a significant indicator for E. The fact that H3K4me3 is significant for EM could be related to its role in transcriptional regulation or it could be due to nucleosome displacement that occurs at regions that are replicating.

An interesting pattern was observed in the epigenetic group that contained H3K27me3 alone. E and EM replicating segments were significantly depleted in the epigenetic groups that were marked by H3K27me3, while M, ML and L were all significantly enriched. However, any group that maintained H3K27me3 in combination with H3K4me3 or H3K56ac lost any significant depletion or enrichment for all RT classes. This supports the role of H3K27me3 as a marker for facultative heterochromatin and again, suggests that its presence does not drive RT. Instead, its notable association with
later replicating sequences seems to be a consequence of other factors contributing to a heterochromatic environment.

*Replication timing is associated with transcriptional activity*

To study the relationship between replication timing and transcription we conducted an RNA-Seq experiment and analyzed transcriptional levels in the replication time classes. To account for sequencing bias across the genome, RNA-Seq data were normalized against a total genomic DNA control. Transcription data was placed into 1-kb bins and the average fold-enrichment (FE) was calculated for each bin. Bins of transcriptional activity were assigned to their appropriate RT class and mean differences between the classes were compared using the Games-Howell test (Figure 8).

We found that genome-wide transcriptional activity is significantly higher in earlier replicating regions than in later replicating regions, suggesting that the transcriptional activity is coordinated with replication timing. Similar results were reported for Drosophila and mammals (Schwaiger, et al. 2009; Hiratani, et al. 2010; Ryba, et al. 2010). Unexpectedly, the highest levels of transcriptional activity were associated with EM and not E replication timing. The ability to observe this subtle difference in transcription demonstrates the improvement of our segmentation method over the use early to late ratios used in previous RT experiments. We showed earlier that the presence of H3K4me3 alone or in combination with H3K56ac is associated with an enrichment of EM. On the other hand, early RT is most highly enriched in regions marked by H3K56ac, which is a mark for euchromatin. Together these results indicate RT is more highly associated with the chromatin state than transcription. Schwaiger, et al. (2009) proposed a similar idea based on the human β-globin gene that replicates early in cells where it is expressed but late in cells where it is not expressed. When gene expression was prevented by removal of the enhancer, the region containing the gene still replicated early. Taken together, these results suggest that H3K56ac contributes to a euchromatic state that is accessible for transcription and replication. Accordingly,
early replicating sequences are associated with H3K56ac alone or in combination with H3K4me3. Although open chromatin state predicted for early RT contributes to accessibility for transcription, H3K4me3 seems to contribute more to transcriptional regulation while H3K56ac is more related to early transcription and perhaps early origin firing.

Conclusions

This is the first genome-wide RT study in Arabidopsis that compares epigenetic and transcriptional profiles in the same cell line under the same growth conditions. Our data suggest that RT is regulated by at the level of higher order chromatin structure. Demonstration that the RT domains coincide with genomic interactions could strengthen this hypothesis. Analysis of combinatorial epigenetic marks that have been linked to chromatin state could also help to further elucidate the relationships between these processes. An epigenetic combinatorial analysis of 28 different histone modifications demonstrated that different combinations of epigenetic marks predicted RT with significant accuracy (Comoglio and Paro 2014). A combinatorial analysis in Arabidopsis revealed that the four chromatin states characterize the genome (Roudier, et al. 2011). It would be interesting to compare a combination of these marks with RT data.

Materials and Methods

Cell growth conditions

The A. thaliana suspension cell line Col-0 (ecotype Columbia) was grown in 250 mL flasks containing 50 mL culture medium [3.2 g/L of Gamborg’s B5 basal medium with minimal organics, 3 mM MES, 3% sucrose, 1.1 mg/L 2,4-D, pH 5.8]. Cell flasks were maintained under constant light at 23°C on a rotary shaker at 160 rpm. Cells were sub-cultured every 7 days by aliquoting 6 mL of the cell suspension into 50 mL fresh culture medium. To maximize the number of cells in the logarithmic phase of growth at the
time of assay, 25 mL of 7-day old cells were transferred to 25 mL of fresh medium and allowed to grow for 16 hours prior to harvesting. The 16 hour time point was chosen based upon a preliminary study which found that BrdU incorporation into DNA is maximized at this time (Lee et al 2010).

Sample preparation for RNA-Seq, ChIP-Seq and Repli-Seq varied beyond the 16 post 7-day split and is described below.

**RNA-Seq sample preparation**

Cells were collected at 16 hours after sub-culturing into fresh medium as described. One flask of cells comprised a single bioreplicate and three bioreplicates were collected. Cells were washed three times with cold 1x phosphate-buffered saline (PBS), snap-frozen with liquid nitrogen and stored at -80°C. Frozen cells were ground in a mortar and pestle with liquid nitrogen. Total RNA was extracted from 10 µg of frozen ground cell material using PureLink Plant RNA Reagent (Ambion). Residual DNA was removed using Turbo DNA-free kit (Ambion), yielding 7.6 µg of RNA (~76% recovery of RNA). Ribosomal RNA (rRNA) was depleted from 4 µg of DNA-free RNA using the Ribo-Zero Magnetic Kit for Plant Leaf material (EpiCentre). The efficiency of rRNA depletion was determined by quantitative real-time polymerase chain reaction (qRT-PCR) using primers for the Arabidopsis 5S, 18S, 23S and 25S rRNAs. qRT-PCR was performed with qScript One-Step SYBR Green qRT-PCR kit, Low ROX (Quanta BioSciences). RNA-seq libraries were prepared with the ScriptSeq™ v2 RNA_seq Library Preparation Kit (EpiCentre). Each sample was uniquely barcoded with ScriptSeq™ Index PCR Primers (EpiCentre) and samples were pooled for sequencing on one lane of an Illumina HiSeq 2000 instrument.

**Epigenetic ChIP-seq assays**

Sixteen hours after a post 7-day splits, cells were crosslinked with 1% paraformaldehyde for 10 minutes at room temperature and washed three times with
cold 1x phosphate buffered saline (PBS). The samples were flash frozen with liquid nitrogen. The frozen cell pellet was ground to a fine powder with mortar and pestle and stored at -80°C. Chromatin immunoprecipitation was performed on the frozen ground cell material according to Gendrel, et al. (2005) with modifications. Chromatin was sonicated using the S2 Covaris with parameters: duty = 10%, intensity = 5, bursts/sec = 200) optimized to obtain a final DNA fragment size of 200-300 bp. The sonicated chromatin was precipitated overnight at 4°C with antibodies to H3K4me3 (Millipore, Billerica, Massachusetts, USA), H3K27me3 (Millipore) or H3K56ac (Millipore). The immunoprecipitates were washed, crosslinks were reversed and samples were purified using the Qiagen PCR Purification kit. The NEXTflex Illumina Chip-Seq Library Prep Kit with the ultra-low input protocol was used to generate sequencing libraries. The libraries were sequenced on an Illumina HiSeq 2000 instrument.

Repli-Seq sample preparation

Labeling and fixation

Cell growth conditions for the repli-seq experiment were the same as described previously. At 16 hours after the 7-day split, cells were labeled with 10 µM 5-Ethynyl-2’-deoxyuridine (EdU, Life Technologies) for 30 minutes and then fixed for 10 minutes in 1% paraformaldehyde at room temperature on a slow rocker. The fixed cells were washed three time with 1X phosphate buffered saline (PBS), snap-frozen in liquid nitrogen and stored at -80°C.

Nuclei extraction and purification

Nuclei were isolated and purified as described by Folta and Kaufmann (2006) with modifications. Frozen cells pellets were ground in 100 mL of cold cell lysis buffer (15 mM Tris, pH 7.5, 2 mM EDTA, 80 mM KCl, 20 mM NaCl, 0.1% Triton X-100, 15 mM β-mercaptoethanol) using a small kitchen blender (Cuisinart SmartPower) at 4°C. The ground cell suspension was incubated for 5 additional minutes and filtered through two
layers of Miracloth. The sample was collected in chilled 50-mL falcon tubes, centrifuged at 400 g for five minutes and the supernatant was discarded. The nuclei pellet was carefully re-suspended and diluted to 30 mL in Folta EB (20 mM PIPES-KOH, pH 7.0, 2M hexylene glycol, 10 mM MgCl2, 5mM β-mercaptoethanol). Percoll solutions were prepared at 80% and 30% in gradient buffer (5 mM PIPES-KOH, pH 7.0, 0.5 M hexylene glycol, 10 mM MgCl2, 1% Triton X-100). The percoll density gradient was prepared in 50-mL round bottom glass centrifuge tubes by layering 6 mL of 80% percoll solution beneath 6 mL of 30% solution. The 30-mL nuclei re-suspension was layered above the 30% percoll layer with care taken to maintain a sharp interface between the layers. The gradients were centrifuged at 1500 g for 30 min. at 4°C. After centrifugation, a thin white band of nuclei was present at the interface between the 30% and 80% percoll layers. The nuclei were removed and transferred to a new 50-mL glass centrifuge tube. The nuclei were re-suspended and diluted to 10 mL in gradient buffer. Six mL of 30% percoll solution were layered underneath the nuclei suspension. The gradient was centrifuged at 1500 g for 10 min. at 4°C. The supernatant was discarded and the remaining nuclei pellet was washed in 3 mL of gradient buffer, transferred to a 15-mL falcon tube and centrifuged at 200 g for 5 min. at 4°C. The supernatant was removed and the nuclei pellet was washing in 3 mL of modified LB and centrifuged at 200 g for 5 min. The supernatant was removed.

AF488 conjugation to EdU and FACS

The Click-it®EdU Alexa fluro-488 kit (Life Technologies) was used to conjugate AF488 to EdU incorporated into the DNA during labeling. The nuclei pellet was re-suspended in 1 mL of Click-it cocktail, rotated in the dark for 30 min. at room temperature and centrifuged at 200 g for 5 min. at 4°C. The supernatant was removed and washed in 3 mL LB as previously described. The remaining nuclei pellet was re-suspended and diluted to 10 mL in LB. The sample was stored overnight in the dark at 4°C. The following day, prior to FACS, the sample was centrifuged as previously described and
the supernatant was removed. The nuclei pellet was re-suspended in LB containing 2 µg/mL DAPI for 5 minutes. To remove nuclei aggregates, the nuclei suspension was filtered on ice using a 20-micron nylon filter. The quality and density of nuclei were assessed under a microscope. Occasionally, additional dilution in the LB-DAPI solution and/or filtering were required to reduce aggregation of the nuclei.

The BD Influx flow cytometer was used to sort and recover early, middle, late S-phase populations of nuclei. The G1 population of nuclei was also recovered to serve as the whole genome control. The cytometer is equipped with UV (535 nm) and blue (488 nm) lasers and 460/50 nm and 530/40 nm emission filters, which allow two-way sorting of nuclei based on amount of DAPI (total DNA) and AF488 conjugated EdU (replicating DNA). Nuclei were sorted into 1.5-mL Lo-Bind tubes containing 100 µL NaCl-Tris-EDTA (STE) and stored at -80°C.

Repli-Seq DNA extraction and AF488 immunoprecipitation

Tubes containing sorted nuclei were thawed and centrifuged at 400 g for 5 minutes. The supernatant was removed. The nuclei pellet was re-suspended in 500 µL reverse crosslinking mixture (EDTA 50 mM, 1% sarkosyl, 200 µg/mL Proteinase K) and incubated at 42°C for one hour. Crosslinks were reversed by incubation at 65°C overnight. DNA was extracted and sonicated using a S2 Covaris with parameters: duty = 10%, intensity = 5, bursts/sec = 200) optimized to obtain a final DNA fragment size of 200-300 bp.

Sequencing library construction

Repli-Seq libraries for were generated with the NEXTflex Illumina Chip-Seq Library Prep Kit (BioScientific) with ultra-low input protocol. In brief, uniquely barcoded Illumina Truseq adapters are ligated to processed DNA fragment and PCR is used to enrich size selected DNA. All libraries were sequenced on an Illumina HiSeq platform.
**Bioinformatic analysis**

**Transcriptome analysis**

Sequencing adapters were removed from raw reads with Trim Galore! FastQC /0.10.1 was used to assess the quality of the reads. The TopHat2 aligner was used to map reads to the Arabidopsis Tair10 genome. Samtools was used to remove redundant and improperly paired mapped reads. MACS2/2.1.0 was used to generate redundant and improperly paired mapped reads. MACS2 callpeak module was used to convert processed RNA-Seq bam files to normalized “signal per million reads” (SPMR) bedgraphs with default settings and the –SPMR flag. Fold enrichment (FE) of the RNA-Seq sample to input control was calculated using the normalized SPMR bedgraphs with the macs2 bdgcmp module at default setting and the following flag: -m FE. The bedtools module was used to bin the Arabidopsis genome into 1-kb windows. The mean FE was calculated for all windows across the genome.

**Epigenetic ChIP-Seq**

ChIP-Seq reads were mapped to the TAIR 10.0 version of the Arabidopsis thaliana genome using the Burrows-Wheeler Alignment package with the bwa-mem algorithm (bwa/0.7.12) (Li and Durbin 2009). The deepTools package was used to assess correlation between biological replicates. After correlation analysis was performed, ChIP bam files of each biological replicate were merged with samtools.

MACS2/2.1.0 was used to analyze ChIP-Seq data (Liu 2014). To call peaks of enrichment and to generate normalized signal per million reads (SPMR) bedgraph files, the macs2 callpeak module was used with default parameters and the following additional flags: --SPMR --broad --broad-cutoff 0.01. Fold enrichment (FE) of the ChIP sample to control was calculated using the normalized SPMR bedgraphs in the macs2 bdgcmp module with default setting and the following flag: -m FE. The bedtools module was used to bin
the Arabidopsis genome into 1-kb windows. The mean FE was calculated for all windows across the genome.

Replication timing

FastQC /0.10.1 was used to assess the quality of the raw sequencing data. Repli-seq reads were trimmed using Trim Galore! to remove Illumina Truseq adapter index sequences used in multiplexing of samples (Krueger 2015). A read pair was discarded from further analysis if it contained a read with less than 40 bp after trimming.

Repli-seq reads were mapped to the TAIR 10.0 version of the Arabidopsis thaliana genome using the Burrows-Wheeler Alignment package with the bwa-mem algorithm (bwa/0.7.12) (Li and Durbin 2009). Samtools was used to remove duplicate and improperly paired reads. The deepTools package was used to assess correlation between biological replicates.

The Repli-seq program was used to assign RT (Zynda, et al. 2016). First bioreplicate bam files for each S-phase population were merged with samtools. Merged bam files for early, mid and late S and G1 are converted to bedgraph files. Bedgraphs are normalized to a 1X coverage. To control for biases in sequencing, a ratio of S-phase to G1 is calculated for 1 kb bins across the genome. For each bin, the highest normalized signal from either early, mid or late population is identified. Each signal is divided by the highest signal. The result of this normalization step is that the highest signal is scaled to 1 and the remaining signal will range from 0-1. An early, mid or late time is assigned as the predominate replication timing class in a 1 kb window if the signal is greater than 0.5. For example, if in a given window the scaled signal are: early =1, mid =0.6, late = 0.3, then the 1 kb window is assigned predominately “early-mid” replicating. On the other hand if, early =1, mid = 0.3, late = 0.3, the window is “early” replicating.
References


Figure 1. Representative FACS profile of Arabidopsis suspension cell nuclei. Sort gates are indicated by the red box. The early, mid and late S and G1 nuclei fractions were collected.
Figure 2. Early, mid and late S-phase sequencing read pile-up normalized against the G1 reference. Merged biorep data for all five Arabidopsis chromosomes are shown. The small black boxes mark the centromere for each chromosome.
Figure 3. Representative IGV tracks displaying relative read density of RT for chromosome III and the segmentation profile.
Figure 4. General features of Arabidopsis RT. Panel (a) displays the RT pattern for each of the five chromosomes. The distribution of RT classes across the genome is represented in the pie chart (b). Panel (c) displays the size distribution of RT classes. The boxes in the plots represent the 25-75th inner quartile range (IQR) with the median indicated by a solid black line. Lower and upper whiskers correspond to 25th quartile – 1.5*IQR and 75th quartile + 1.5*IQR, respectively.
Figure 5. Representative IGV tracks displaying relative read density of RT for chromosome II and the segmentation profile. The zoom shows an atypical region of late replication (red) in the short of chromosome II corresponding to the NOR.
Figure 6. Average epigenetic FE level across RT classes. The dotted black line indicates IP:control ratio = 1 to delineate enrichment or depletion of epigenetic signal.
Figure 7. Distribution of RT across epigenetic states. The control panel represents the distribution of RT classes across the genome.
Figure 8. Bar plot of the mean level of transcription across RT classes. Groups marked with the same letter are not significantly different.
Table 1. Mappable read counts from Illumina paired-end data.

<table>
<thead>
<tr>
<th>Biorep</th>
<th>Mappable reads</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RNA Seq</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2163613</td>
</tr>
<tr>
<td>2</td>
<td>3388255</td>
</tr>
<tr>
<td>3</td>
<td>6411363</td>
</tr>
<tr>
<td><strong>H3K4me3 ChIP Seq</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>42647765</td>
</tr>
<tr>
<td>2</td>
<td>37944687</td>
</tr>
<tr>
<td>3</td>
<td>37315965</td>
</tr>
<tr>
<td>4</td>
<td>40730119</td>
</tr>
<tr>
<td>Input</td>
<td>29224111</td>
</tr>
<tr>
<td><strong>H3K27me3 ChIP Seq</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>47585034</td>
</tr>
<tr>
<td>2</td>
<td>44386074</td>
</tr>
<tr>
<td>3</td>
<td>33817532</td>
</tr>
<tr>
<td>4</td>
<td>52698114</td>
</tr>
<tr>
<td>Input</td>
<td>43856509</td>
</tr>
<tr>
<td><strong>H3K56ac ChIP Seq</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>41771588</td>
</tr>
<tr>
<td>2</td>
<td>36209233</td>
</tr>
<tr>
<td>3</td>
<td>37171210</td>
</tr>
<tr>
<td>4</td>
<td>30986616</td>
</tr>
<tr>
<td>Input</td>
<td>41976627</td>
</tr>
<tr>
<td><strong>Repli-Seq</strong></td>
<td></td>
</tr>
<tr>
<td>Early 1</td>
<td>6541753</td>
</tr>
<tr>
<td>Early 2</td>
<td>3335710</td>
</tr>
<tr>
<td>Early 3</td>
<td>9244254</td>
</tr>
<tr>
<td>Mid 1</td>
<td>12957879</td>
</tr>
<tr>
<td>Mid 2</td>
<td>7022280</td>
</tr>
<tr>
<td>Mid 3</td>
<td>7044808</td>
</tr>
<tr>
<td>Late 1</td>
<td>5256721</td>
</tr>
<tr>
<td>Late 2</td>
<td>6212639</td>
</tr>
<tr>
<td>Late 3</td>
<td>2697429</td>
</tr>
<tr>
<td>G1 Input</td>
<td>12204470</td>
</tr>
</tbody>
</table>
Table 2. Description of the possible epigenetic signatures for 1-kb windows across the genome. The “-” and “+” in each column indicates a signature defined as depleted or enriched, respectively, for that epigenetic mark.

<table>
<thead>
<tr>
<th>Epigenetic signature</th>
<th>H3K4me3</th>
<th>H3K27me3</th>
<th>H3K56ac</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K56ac</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>H3K56ac + H3H4me3</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>H3K4me3 + H3K27me3</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>H3K4me3 + H3K27me3 + H3K56ac</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>H3K27me3+H3K56ac</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 3. Chi-square table comparing the distribution of RT across the epigenetic signatures. The bold font indicates that a given RT class and epigenetic signature occurred together more often than expected by chance.

<table>
<thead>
<tr>
<th>Time class</th>
<th>K56</th>
<th>K56 K4</th>
<th>K4</th>
<th>K4 K27</th>
<th>K4 K27 K56</th>
<th>K27 K56</th>
<th>K27</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>observed</td>
<td>2613</td>
<td>6544</td>
<td>888</td>
<td>244</td>
<td>609</td>
<td>190</td>
<td>3053</td>
<td>14141</td>
</tr>
<tr>
<td>expected</td>
<td>1468</td>
<td>5962</td>
<td>1013</td>
<td>217</td>
<td>493</td>
<td>144</td>
<td>4845</td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>892.44</td>
<td>56.89</td>
<td>15.32</td>
<td>3.46</td>
<td>27.35</td>
<td>14.49</td>
<td>662.63</td>
<td></td>
</tr>
<tr>
<td>$p$-value</td>
<td>5.81E-17</td>
<td>1.7E-4</td>
<td>0.91</td>
<td>1</td>
<td>0.29</td>
<td>0.93</td>
<td>1.77E-124</td>
<td></td>
</tr>
<tr>
<td><strong>ES MS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>observed</td>
<td>3384</td>
<td>16910</td>
<td>2955</td>
<td>494</td>
<td>1036</td>
<td>271</td>
<td>8541</td>
<td>33591</td>
</tr>
<tr>
<td>expected</td>
<td>3488</td>
<td>14162</td>
<td>2405</td>
<td>515</td>
<td>1171</td>
<td>343</td>
<td>11508</td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>3.09</td>
<td>533.44</td>
<td>125.68</td>
<td>0.83</td>
<td>15.53</td>
<td>15.01</td>
<td>765.10</td>
<td></td>
</tr>
<tr>
<td>$p$-value</td>
<td>1</td>
<td>1.85e-97</td>
<td>9.33e-16</td>
<td>1</td>
<td>0.90</td>
<td>0.92</td>
<td>4.81e-146</td>
<td></td>
</tr>
<tr>
<td><strong>MS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>observed</td>
<td>319</td>
<td>1729</td>
<td>380</td>
<td>121</td>
<td>252</td>
<td>71</td>
<td>4588</td>
<td>7460</td>
</tr>
<tr>
<td>expected</td>
<td>775</td>
<td>3145</td>
<td>534</td>
<td>114</td>
<td>260</td>
<td>76</td>
<td>2556</td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>267.96</td>
<td>637.56</td>
<td>44.49</td>
<td>0.39</td>
<td>0.25</td>
<td>0.34</td>
<td>1615.86</td>
<td></td>
</tr>
<tr>
<td>$p$-value</td>
<td>4.43e-43</td>
<td>3.22e-119</td>
<td>6.70e-3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5.52e-249</td>
<td></td>
</tr>
<tr>
<td><strong>MS LS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>observed</td>
<td>535</td>
<td>2697</td>
<td>497</td>
<td>150</td>
<td>389</td>
<td>137</td>
<td>5415</td>
<td>9820</td>
</tr>
<tr>
<td>expected</td>
<td>1020</td>
<td>4140</td>
<td>703</td>
<td>150</td>
<td>342</td>
<td>100.195</td>
<td>3364</td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>230.35</td>
<td>502.94</td>
<td>60.43</td>
<td>0.00</td>
<td>6.38</td>
<td>13.52</td>
<td>1249.93</td>
<td></td>
</tr>
<tr>
<td>$p$-value</td>
<td>1.25e-35</td>
<td>4.08e-91</td>
<td>5.55e-05</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5.52e-249</td>
<td></td>
</tr>
<tr>
<td><strong>LS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>observed</td>
<td>69</td>
<td>217</td>
<td>52</td>
<td>12</td>
<td>37</td>
<td>11</td>
<td>1236</td>
<td>1634</td>
</tr>
<tr>
<td>expected</td>
<td>170</td>
<td>689</td>
<td>117</td>
<td>25</td>
<td>57</td>
<td>17</td>
<td>560</td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>59.72</td>
<td>323.23</td>
<td>36.11</td>
<td>6.79</td>
<td>6.99</td>
<td>1.93</td>
<td>816.76</td>
<td></td>
</tr>
<tr>
<td>$p$-value</td>
<td>6.99e-05</td>
<td>3.42e-54</td>
<td>0.05</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5.95e-157</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>observed</td>
<td>6920</td>
<td>28097</td>
<td>4772</td>
<td>1021</td>
<td>2323</td>
<td>680</td>
<td>22833</td>
<td>66646</td>
</tr>
</tbody>
</table>
Chapter IV

DNA replication timing displays plasticity in the pericentromere during dark adaptation of Arabidopsis

Introduction

DNA replication timing (RT) refers to the temporal pattern of DNA duplication during S-phase. In mice, humans and Drosophila, the RT program is highly conserved across developmentally related cells, so that specific regions of chromosome replicate at well defined times during S-phase (Ryba, et al. 2010). In contrast, studies of cells of different developmental lineages have shown that changes in the RT program occur during development (Schwaiger, et al. 2009; Farkash-Amar and Simon 2010; Rhind and Gilbert 2013). As much as 20-50% of the genome undergoes a switch in RT when mammalian or Drosophila cells from different lineages are compared (Schwaiger, et al. 2009; Hansen, et al. 2010; Hiratani, et al. 2010).

Changes in RT are often accompanied by developmentally related changes in gene expression (Schubeler, et al. 2002; Woodfine, et al. 2004). For example, early to late RT changes between murine embryonic stem cells and epiblast cells are associated with repression of mESC-specific genes (Hiratani, et al. 2010). Furthermore, late to early transitions were associated with activation of germ layer-specific (e.g. post-epiblast stage) genes. Replication timing has also been linked to epigenetic marks that impact gene expression and chromatin structure. For example, increased acetylation of histone lysine residues is associated with early replication and euchromatin while decreased acetylation is associated with late replication and more highly condensed chromatin (Azuara, et al. 2006; Schwaiger, et al. 2009; Lee, et al. 2010; Casas-Delucchi, et al. 2012). Late replication is associated with enrichment of H3K27me3, a repressive epigenetic mark that is regulated by the polycomb group (PcG) proteins that are involved in developmental gene expression (Eaton, et al. 2011; Lo Sardo, et al. 2013).
Previously we established that the epigenetic state and transcription are related to genome-wide RT patterns in light grown Arabidopsis thaliana suspension cell culture (Chapter 3). The epigenetic marks, H3K4me3 and H3K56ac, are associated with earlier replication but are significantly depleted in sequences that replicate at later times in S-phase. In contrast, H3K27me3 is depleted in early replicating sequences but is enriched at later regions that replicated. We also found that transcriptional levels are higher at earlier replicating regions than regions that replicate later. These findings support co-regulation of RT with epigenome and transcription but to elucidate the functional interplay between RT, epigenetics and transcription, it necessary to study how they change in relation to one another under different biological/experimental conditions.

The response to light in plants elicits shifts in chromatin dynamics (Tessadori, et al. 2007; Bourbousse, et al. 2015). Changes in light intensity and spectral quality (e.g. red to far-red ratio) can change expression levels in up 30% of Arabidopsis genes, indicating that light responsiveness has a strong impact on transcription (van Zanten, et al. 2012). Furthermore, chromatin becomes less compact in the nucleus of plants grown under lower light intensities or under conditions that mimic shading in nature (Tessadori, et al. 2009; van Zanten, et al. 2010). Guo, et al. (2008) conducted a chromatin immunoprecipitation (ChIP) experiment to study changes in the levels of the active epigenetic marks, H3K9ac and H3K4me3, and the repressive marks, H3K9me2 and H3K27me3, in seedlings grown under light and dark conditions. They found that the levels of these marks differed at several light-regulated genes, which supports a role for epigenetic regulation of transcription during the light response.

In an earlier study we compared genome-wide changes in H3K4me3 and H3K27me3 levels at differentially expressed genes between light and dark-adapted cell lines. Overall, there was a weak association between the changes in gene expression and epigenetic enrichment as determined by the MACS2 peak caller. However, we demonstrated that the overlapping approach for analyzing the different data types at
the genome-wide level can be misleading because differences in quantitative levels of epigenetic enrichment are not accounted for in peak calling.

We extended our earlier experiments to study the RT changes that occur during dark-adaptation of Arabidopsis cells. We found that 23% of the genome switches RT between the light and dark conditions. Many regions that changed RT were small (1 kb) and were interspersed throughout the chromosome arms. There is a weak correlation between RT changes with transcriptional and epigenetic changes at the genome-wide level. A notable exception to the small, dispersed RT changes is that large regions of localized change were observed primarily in or near the pericentromere.

This is the first reported study of genome-wide changes in a RT program in plants. Furthermore, while other eukaryotic RT studies compare developmental or across species changes, we used a genetically identical cell line that differed only in the treatment. Thus, we were able to study the effect of environmental influence on RT. Ultimately, our work will help to elucidate the role of chromatin dynamics in response to environmental changes.

**Results**

*General features of replication timing Light and Dark-adapted suspension cells*

To generate genome wide replication timing profiles of Arabidopsis, light and dark grown cells were pulse labeled for 30 minutes with 5-Ethynyl-2'-deoxyuridine (EdU) as previously described (Wear et al. 2016). Nuclei were isolated and separated into early, mid and late S-phase by FACS. EdU-labeled DNA from each S-phase sub-population was immunoprecipitated, sequenced and mapped against the TAIR10 Arabidopsis genome. Table 1 shows the mappable read counts obtained for S-phase fractions and G1-phase. Sequencing data of bioreplicates were merged for the early, mid and late populations, normalized to the G1 background reference and smoothed (Li et al. 2009; Zynda et al 2016). A novel algorithm was used to assess enrichment of sequences from each
population, and every 1-kb region of the genome was assigned a replication timing class of “Early” (E), “Early-Mid” (EM), “Mid” (M), “Mid-Late” (ML), “Late” (L), Early-Late” (EL) or “Early-Mid-Late” (EML). Regions of the genome that were indeterminate (i.e. low sequencing coverage) were assigned as “NA” (Zynda et al. 2016).

The top panel of Figure 1 displays the chromosome III RT profiles for early, mid and late S-phase for the light and dark samples. Early sequences are enriched in the euchromatic arms while late replicating sequences are enriched in the centromere and pericentromere regions. Mid-S sequences were evenly distributed across the chromosome with a slight depletion near the centromere. The lower panel of Figure 1 contains RT segmentation tracks that were generated from the early, mid and late profiles. At this level of resolution the light and dark RT segmentation patterns were highly similar. However, there were notable differences in the RT profiles in the pericentromeric region where several segments of ML (yellow) in light became L (red) in dark.

To gain greater insight into the genome-wide changes, we first compared the distribution of RT classes for light and dark (Figure 2a,b). As we observed for chromosome III, the distribution of RT classes is mostly similar in the two treatments. Most of the genome replicates in E and EM. However, combining the total genome amount of these two RT classes, there is less E and EM in dark (66%) than in light (69%). There is a slight increase in the amount of M and L in the dark compared to the light but the amount of ML is lower. It is difficult to ascertain whether the increased M and L observed in the dark were related to the decrease in ML or E. Nonetheless, when the genome proportions for the classes are combined, the dark is more enriched for M, ML and L (32%) than the light (29%).

The light and dark genome contained the same amount of EL (0%) and EML (2%). These RT classes were excluded from further analyses for two reasons. First, a very small fraction of the genome was represented by these timing classes, which makes it
difficult to test assumptions about their function. Second, the direction of RT changes to or from these classes is ambiguous. For example, a change from M to EL during dark transition could be designated as “earlier” or “later” or both.

Minor differences in RT segment sizes were observed for light and dark (Figure 2 c,d). Most RT segments are larger in light, with the exception of the late replicating segments. The smaller segment sizes observed for most of the timing classes in dark could indicate that RT transitions occurred more often in dark than in the light. A greater number of transitions in RT in the dark could be related to a stress or adaptive response. This idea could be tested by comparing RT changes under treatments such as drought or nutrient deprivation in whole plants. Increases in the amount of L replication and size of the L segments in the dark relative to light could reflect a protective mechanism to prevent activation of transposons that reside in or near the centromere. Based on the comparison, we concluded that the dark adaptation impacts RT in Arabidopsis cell culture.

Differences in replication timing

An important goal of this work was to understand how RT is altered in a genetically identical cell line that differs only in a treatment. The experimental system is unique in that other genome-wide studies of RT compare changes that occur between cells in various stages of lineage specification or in cells at the same development stage across different species.

In our Arabidopsis cells, we calculated that 77% of genome replicated at the same time while 23% of the genome was comprised of regions that changed their time of replication between the two treatments. These regions will henceforth referred to as regions of altered timing (RATs). The 23% change in RT is considerably less compared to changes that occur during mammalian development, which is usually ~50% of the genome (Woodfine, et al. 2004; Hiratani, et al. 2010; Rivera-Mulia and Gilbert 2016). Regions of altered timing were grouped based upon 1) directionality of the RT change
from light to dark - i.e. “earlier-to-later” vs “later-to-earlier” - and 2) degree of timing change i.e. half step vs. full step. The degree of timing change can be described by considering the RT classes in a linear manner (E <-> EM <-> M <-> ML <-> L). A half step change refers to shift from one timing class to a timing class that is adjacent to it in either direction. For example, E to EM is a half step in either direction of Earlier-to-later or Later-to-earlier. A full step is a change from one timing class to any other timing class that is not adjacent to it. Thus a full step characterizes changes such as E to M, E to ML and E to L. A shift in RT from EM to ML is also a full step since those classes are not adjacent when placed on a linear scale. This resulted in four possible group placements for a single RAT (Table 2). To assist in visualization of the degree and direction of change, RATs are color coded as follows: Earlier-to-later, half step = light blue; Earlier-to-later, full step = dark blue; Later-to-earlier, half step = light red; Later-to-earlier, full step = dark red. An IGV image of the light and dark RT segmentation with their corresponding RATs beneath is shown for chromosome III (Figure 3). There is a noticeable absence of RATs directly in the centromeric region, which is dominated by L replication. RAT density was greatest in the pericentromeres and decreased moving towards distal arms. A surprising observation was the largest RATs for each chromosome are found within the in pericentromeric regions (data not shown).

The largest and most abundant RATs are Earlier-to-later half steps (Figure 4) followed by Later-to-earlier half steps. Full steps in either direction are smaller and less common. It is difficult to compare these changes to those derived from other RT experiments. Most studies in eukaryotes designate a RT switch by comparing profiles that are characterized by either E or L timing classes. In our experiment, five possible RT classifications can change in two different manners (i.e. direction and degree). The distribution of RATs supports the visual assessment that replication timing between light and dark is mostly conserved. However, we can conclude that when there is a RT switch during dark adaptation, the change is usually subtle, as suggested by the high number of half steps, and that the change is most frequently a shift to a later replication
time in the dark-grown as compared to light-grown cells. It is interesting to note that earlier-to-later changes are associated with loss of pluripotency during murine embryogenesis (Hiratani, et al. 2010).

**Transcription and epigenetics at RATs**

Replication timing is associated with the epigenome, and changes in RT are sometimes accompanied by changes in the epigenome. To determine if RATs display corresponding changes in the epigenetic levels, we calculated the mean fold-enrichment (FE) of epigenetic sequencing signal to the reference signal in 1-kb bins across the genome. We then plotted the light FE against the dark FE for every bin that comprised a RAT (Figure 5). The scatter plots revealed a near linear relationship between light and dark for epigenetic and transcriptional levels at each class of RATs. This indicated that light and dark treatments maintained similar levels of epigenetic enrichment and transcription in regions of altered replication timing.

However, the plots also revealed the existence of outliers (e.g. RATs with large differences in either epigenetic level or transcription). This led to a second approach to analyze the RAT data. To increase stringency, we focused on the strongest changes in RT, which corresponded to full step RATs. For each 1-kb window that comprised a RAT, the difference in FE between light and dark was calculated for each epigenetic mark and transcription. Figure 6 displays the distribution of the differences and the corresponding box and whiskers plot. An outlier was defined as a value that was above or below 1.5*IQR of the data. In other words, an outlier is a RAT with a large difference in epigenetic and/or transcription level between light and dark grown cells. Outliers are indicated by the black dots in Figure 6. We isolated the “outlier RATs” and viewed them in IGV (Figure 7). With the exception of chromosome 1, most of the outlier RATs are in or very near the pericentromeric region. There is a broader distribution of outlier RATs across chromosome 1 compared to the other chromosomes, but they were absent in the
distal arms as also observed for the other chromosomes. Taken together, the dark treatment primarily impacted outlier RATs in the centromere and pericentromere.

**Discussion**

The goal of this research was to study regulation of DNA replication timing in Arabidopsis thaliana. This is the first genome wide study in plants that compares RT of plant cells grown under different environmental conditions. Replication timing was profiled in light- and dark-adapted cells. An important aspect of this study is that it was conducted in a genetically identical suspension cell culture line, enabling the exclusion of developmental triggers and effects that are present in multi-tissue samples. In this way, we were able to focus primarily on the treatment effect. We also conducted an RNA-Seq experiment and epigenetic ChIP-Seq experiments to compare changes in RT to changes in gene expression and the epigenome.

Regions of altered timing vary in size from 1-kb to 80-kb in length and are distributed across all chromosomes. The largest RAT on each of the Arabidopsis chromosomes is located in or near the centromere region. RATs were grouped based upon direction of change (earlier-to-later vs later-to-earlier) and degree of change (half-step vs full step) and mean transcriptional levels and epigenetic enrichment levels were compared for each RAT group between the light and dark samples. Transcriptional and epigenetic enrichment levels are not significantly different between light and dark samples for any of the RAT groups. However, there are RATs that contain high levels of differences in transcription and epigenetic levels (outliers). Interestingly, we found that RATs that contain epigenetic or transcriptional outliers are distributed primarily in or near the pericentromeric regions for all chromosomes. The exception to this are the outlier RATs on chromosome I, which are more evenly distributed across the short arm. This suggests that RT changes in the centromere and pericentromere are more closely associated with epigenetic or transcriptional features than RATs that are in the euchromatic arms. It also suggests that RATs in these regions are more likely to be
accompanied by changes in transcriptional levels and/or epigenetic enrichment of H3K4me3, H3K27me3 or H3K56ac.

There is evidence that stress can induce other epigenetic changes in the pericentromere. For example, under drought stress, phosphorylation of H3T3 (threonine 3 of histone 3) increases in the pericentromeric region of Arabidopsis (Wang, et al. 2015). It is possible that the H3T3ph epigenetic mark plays a role in chromatin dynamics that regulate RT in this region. Together, our data suggest that the centromere and pericentromere display a plastic response to the dark adaptation period while the largely euchromatic regions are relatively stable. The centromere, comprised of highly repetitive DNA, is a chromosomal region that is directly involved in formation of the kinetochore (Simon, et al. 2015). It is flanked by pericentromeric DNA that is also comprised of highly repetitive DNA, as well as, several silenced transposons.

Although a strict systemic analysis has not yet been performed, visual assessment of the RT changes in the pericentromere seem to correspond to Earlier-to-later changes reflecting primarily ML to L changes in the light to dark adaptation. It is interesting to consider why this region of the genome maintains the vast majority of RT changes towards a later time. Transposon density is high in the pericentromere and centromere in Arabidopsis and their activation can be dangerous for the genome (Grandbastien 2004). Genomic stress can lead to transposon activation (Santos, et al. 2015). One hypothesis is that the constant dark treatment caused a genomic stress, which resulted in increased compaction of the pericentromere activated by increased silencing of DNA to prevent transposon activation (Grandbastien 2004). It is possible that the increased heterochromatin compaction caused a change in RT that was detected in our experiment. Determination of changes in levels of constitutive heterochromatin marks, such as H3K9me2 or DNA methylation, could provide support for the hypothesis that RT changes are driven by changes to the chromatin structure in our experimental
system. For example, higher levels of H3K9me2 in the dark at sequences that replicate L would indicate a more highly condensed heterochromatin in the dark.

The fact that the functional centromere is comprised of L RT under both environmental conditions precludes our ability to draw direct conclusions about potential stress impacts in this region. However, analysis of H3K9me2 or functional studies of the dynamics of centromeric histone variant, CENH3, would better address questions relating to chromatin conformation in the centromere during dark-adaptation.

**Materials and Methods**

**Cell growth conditions**

The *A. thaliana* suspension cell line Col-0 (ecotype Columbia) was grown in 250 mL flasks containing 50 mL culture medium [3.2 g/L of Gamborg’s B5 basal medium with minimal organics, 3 mM MES, 3% sucrose, 1.1 mg/L 2,4-D, pH 5.8]. Cell flasks were maintained under constant light at 23°C on a rotary shaker at 160 rpm. Cells were sub-cultured every 7 days by aliquoting 6 mL of the cell suspension into 50 mL fresh culture medium. Twenty-five milliliters of 7-day old cells were transferred to 25 mL of fresh medium and allowed to grow for 16 hours prior to harvesting. The 16 hour time point was chosen based upon a preliminary study which found that BrdU incorporation into DNA is maximized at this time (Lee et al. 2010).

For the dark treatment, all conditions were the same as described above with the exception that cells were grown in constant dark for two weeks. The dark treatment was applied by growing cells in flasks that were completely covered with aluminum foil. At day 7, a 6 mL aliquot of the dark grown cells was rapidly transferred under low light conditions to fresh medium and maintained in the dark for a second week of adaptation. After seven days, 25 mL of dark-adapted cells culture was placed into 25 mL of fresh medium and the cells were grown in dark for 16 hours. Sample preparation for RNA-
Seq, ChIP-Seq and Repli-Seq varied beyond the 16 post 7-day split and is described below.

**RNA-Seq sample preparation**

Cells were collected at 16 hours after sub-culturing into fresh medium as described. One flask of cells comprised a single bioreplicate and three bioreplicates were collected. Cells were washed three times with cold 1x phosphate-buffered saline (PBS), snap-frozen with liquid nitrogen and stored at -80°C. Frozen cells were ground in a mortar and pestle with liquid nitrogen. Total RNA was extracted from 10 µg of frozen ground cell material using PureLink Plant RNA Reagent (Ambion). Residual DNA was removed using Turbo DNA-free kit (Ambion), yielding 7.6 µg of RNA (~76% recovery of RNA). Ribosomal RNA (rRNA) was depleted from 4 µg of DNA-free RNA using the Ribo-Zero Magnetic Kit for Plant Leaf material (EpiCentre). The efficiency of rRNA depletion was determined by quantitative real-time polymerase chain reaction (qRT-PCR) using primers for the Arabidopsis 5S, 18S, 23S and 25S rRNAs. qRT-PCR was performed with qScript One-Step SYBR Green qRT-PCR kit, Low ROX (Quanta BioSciences). RNA-seq libraries were prepared with the ScriptSeq™ v2 RNA_seq Library Preparation Kit (EpiCentre). Each sample was uniquely barcoded with ScriptSeq™ Index PCR Primers (EpiCentre) and samples were pooled for sequencing on one lane of an Illumina HiSeq 2000 instrument.

**Epigenetic ChIP-seq assays**

Sixteen hours after sub-culturing into fresh medium, cells were crosslinked with 1% paraformaldehyde for 10 minutes at room temperature and washed three times with cold 1x phosphate buffered saline (PBS). The samples were flash frozen with liquid nitrogen. The frozen cell pellet was ground to a fine powder with mortar and pestle and stored at -80°C. Chromatin immunoprecipitation was performed on the frozen ground cell material according to Gendrel, et al. (2005) with modifications. Chromatin was sonicated using the S2 Covaris with parameters: duty = 10%, intensity = 5, bursts/sec =
200) optimized to obtain a final DNA fragment size of 200-300 bp. The sonicated chromatin was precipitated overnight at 4°C with either H3K4me3 (Millipore, Billerica, Massachusetts, USA), H3K27me3 (Millipore) or H3K56ac (Millipore). The immunoprecipitates were washed, crosslinks were reversed and samples were purified using the Qiagen PCR Purification kit. The NEXTflex Illumina Chip-Seq Library Prep Kit with the ultra-low input protocol was used to generate sequencing libraries. The libraries were sequenced on an Illumina HiSeq 2000 instrument.

**Repli-Seq sample preparation**

**Labeling and fixation**

Cell growth conditions for the repli-seq experiment were the same as described previously. At 16 hours after the 7-day split, cells were labeled with 10 µM 5-Ethynyl-2’-deoxyuridine (EdU, Life Technologies) for 30 minutes and then fixed for 10 minutes in 1% paraformaldehyde at room temperature on a slow rocker. The fixed cells were washed three times with 1X phosphate buffered saline (PBS), snap-frozen in liquid nitrogen and stored at -80°C.

**Nuclei extraction and purification**

Nuclei were isolated and purified as described by Folta and Kaufmann (2006) with modifications. Frozen cells pellets were ground in 100 mL of cold cell lysis buffer (15 mM Tris, pH 7.5, 2 mM EDTA, 80 mM KCl, 20 mM NaCl, 0.1% Triton X-100, 15 mM β-mercaptoethanol) using a small kitchen blender (Cuisinart SmartPower) at 4°C. The ground cell suspension was incubated for 5 additional minutes and filtered through two layers of miracloth. The sample was collected in chilled 50-mL falcon tubes, centrifuged at 400 g for five minutes and the supernatant was discarded. The nuclei pellet was carefully re-suspended and diluted to 30 mL in Folta EB (20 mM PIPES-KOH, pH 7.0, 2M hexylene glycol, 10 mM MgCl2, 5mM β-mercaptoethanol). Percoll solutions were prepared at 80% and 30% in gradient buffer (5 mM PIPES-KOH, pH 7.0, 0.5 M hexylene...
glycol, 10 mM MgCl2, 1% Triton X-100). The percoll density gradient was prepared in 50-mL round bottom glass centrifuge tubes by layering 6 mL of 80% percoll solution beneath 6 mL of 30% solution. The 30-mL nuclei re-suspension was layered above the 30% percoll layer with care taken to maintain a sharp interface between the layers. The gradients were centrifuged at 1500 g for 30 min. at 4°C. After centrifugation, a thin white band of nuclei was present at the interface between the 30% and 80% percoll layers. The nuclei were removed and transferred to a new 50-mL glass centrifuge tube. The nuclei were re-suspended and diluted to 10 mL in gradient buffer. Six mL of 30% percoll solution were layered underneath the nuclei suspension. The gradient was centrifuged at 1500 g for 10 min. at 4°C. The supernatant was discarded and the remaining nuclei pellet was washed in 3 mL of gradient buffer, transferred to a 15-mL falcon tube and centrifuged at 200 g for 5 min. at 4°C. The supernatant was removed and the nuclei pellet was washing in 3 mL of modified LB and centrifuged at 200 g for 5 min. The supernatant was removed.

AF488 conjugation to EdU and FACS

The Click-iT®EdU Alexa fluor-488 kit (Life Technologies) was used to conjugate AF488 to EdU incorporated into the DNA during labeling. The nuclei pellet was re-suspended in 1 mL of Click-iT cocktail, rotated in the dark for 30 min. at room temperature and centrifuged at 200 g for 5 min. at 4°C. The supernatant was removed and the pellet washed in 3 mL LB as previously described. The remaining nuclei pellet was re-suspended and diluted to 10 mL in LB. The sample was stored overnight in the dark at 4°C. The following day, prior to FACS, the sample was centrifuged as previously described and the supernatant was removed. The nuclei pellet was re-suspended in LB containing 2 µg/mL DAPI for 5 minutes. To remove nuclei aggregates, the nuclei suspension was filtered on ice using a 20-micron nylon filter. The quality and density of nuclei were assessed under a microscope. Occasionally, additional dilution in the LB-DAPI solution and/or filtering were required to reduce aggregation of the nuclei.
The BD Influx flow cytometer was used to sort and recover early, middle, late S-phase populations of nuclei. The G1 population of nuclei was also recovered to serve as the whole genome control. The cytometer is equipped with UV (535 nm) and blue (488 nm) lasers and 460/50 nm and 530/40 nm emission filters, which allow two-way sorting of nuclei based on amount of DAPI (total DNA) and AF488 conjugated EdU (replicating DNA). Nuclei were sorted into 1.5-mL Lo-Bind tubes containing 100 µL NaCl-Tris-EDTA (STE) and stored at -80°C.

Repli-Seq DNA extraction and AF488 immunoprecipitation

Tubes containing sorted nuclei were thawed and centrifuged at 400 g for 5 minutes. The supernatant was removed. The nuclei pellet was re-suspended in 500 µL reverse crosslinking mixture (EDTA 50 mM, 1% sarkosyl, 200 µg/mL Proteinase K) and incubated at 42°C for one hour. Crosslinks were reversed by incubation at 65°C overnight. DNA was extracted and sonicated using a S2 Covaris with parameters: duty = 10%, intensity = 5, bursts/sec = 200) optimized to obtain a final DNA fragment size of 200-300 bp.

Sequencing library construction

Repli-Seq libraries for were generated with the NEXTflex Illumina Chip-Seq Library Prep Kit (BioScientific) with ultra-low input protocol. In brief, uniquely barcoded Illumina Truseq adapters are ligated to processed DNA fragment and PCR is used to enrich size selected DNA. All libraries were sequenced on an Illumina HiSeq platform.

Bioinformatic analysis

Transcriptome analysis

Sequencing adapters were removed from raw reads with Trim Galore! FastQC /0.10.1 was used to assess the quality of the reads. The TopHat2 aligner was used to map reads to the Arabidopsis Tair10 genome. Samtools was used to remove redundant and
improperly paired mapped reads. MACS2/2.1.0 was used to generate transcriptome data. The MACS2 callpeak module was used to convert processed RNA-Seq bam files to normalized “signal per million reads” (SPMR) bedgraphs with default settings and the --SPMR flag. Fold enrichment (FE) of the RNA-Seq sequencing reads to total genomic DNA reference sequencing reads was calculated using the normalized SPMR bedgraphs with the macs2 bdgcmp module at default setting and the following flag: -m FE. The bedtools module was used to bin the Arabidopsis genome into 1-kb windows. The mean FE of RNA-Seq signal relative to the genomic DNA reference was calculated for all windows across the genome.

**Epigenetic ChIP-Seq**

ChIP-Seq reads were mapped to the TAIR 10.0 version of the Arabidopsis thaliana genome using the Burrows-Wheeler Alignment package with the bwa-mem algorithm (bwa/0.7.12) (Li and Durbin 2009). The deepTools package was used to assess correlation between biological replicates. After correlation analysis was performed, ChIP bam files of each biological replicate were merged with samtools.

MACS2/2.1.0 was used to analyze ChIP-Seq data (Liu 2014). To call peaks of enrichment and to generate normalized signal per million reads (SPMR) bedgraph files, the macs2 callpeak module was used with default parameters and the following additional flags: --SPMR --broad --broad-cutoff 0.01. Fold enrichment (FE) of the ChIP sample to the genomic DNA reference was calculated using the normalized SPMR bedgraphs in the macs2 bdgcmp module with default setting and the following flag: -m FE. The bedtools module was used to bin the Arabidopsis genome into 1-kb windows. The mean FE of the epigenetic ChIP-Seq signal to the genomic DNA reference was calculated for all windows across the genome.
**Replication timing**

FastQC /0.10.1 was used to assess the quality of the raw sequencing data. Repli-seq reads were trimmed using Trim Galore! to remove Illumina Truseq adapter index sequences used in multiplexing of samples (Krueger 2015). A read pair was discarded from further analysis if it contained a read with less than 40 bp after trimming.

Repli-seq reads were mapped to the TAIR 10.0 version of the Arabidopsis thaliana genome using the Burrows-Wheeler Alignment package with the bwa-mem algorithm (bwa/0.7.12) (Li and Durbin 2009). Samtools was used to remove duplicate and improperly paired reads. The deepTools package was used to assess correlation between biological replicates.

The Repli-seq program was used to assign RT (Zynda, et al. 2016). First bioreplicate bam files for each S-phase population were merged with samtools. Merged bam files for early, mid and late S and G1 are converted to bedgraph files. Bedgraphs are normalized to a 1X coverage. To control for biases in sequencing, a ratio of S-phase to G1 is calculated for 1 kb bins across the genome. For each bin, the highest normalized signal from either early, mid or late population is identified. Each signal is divided by the highest signal. The result of this normalization step is that the highest signal is scaled to 1 and the remaining signal will range from 0-1. An early, mid or late time is assigned as the predominate replication timing class in a 1 kb window if the signal is greater than 0.5. For example, if in a given window the scaled signal are: early =1, mid =0.6, late = 0.3, then the 1 kb window is assigned predominately “early-mid” replicating. On the other hand if, early =1, mid = 0.3, late = 0.3, the window is “early” replicating.
References


Figure 1. Representative IGV tracks displaying relative read density of RT for chromosome III and segmentation profiles.
Figure 2. General features of RT distribution for light and dark. Pie chart displaying the distribution of RT classes across the genome for light (a) and dark (b) grown cells. Box and whisker plot of distribution of RT segment sizes for light (c) and dark (d). Boxes in the plots represent the 25-75th inner quartile range (IQR) with the median indicated by a solid black line. Lower and upper whiskers correspond to 25th quartile – 1.5*IQR and 75th quartile + 1.5*IQR, respectively.
Figure 3. IGV image of chromosome III RT segmentation for light and dark samples with corresponding RATs below. The RAT track is color coded to indicate direction and degree of RT changes as follows: Earlier-to-later, half step = light blue; Earlier-to_later, full step = dark blue; Later-to-earlier, half step = light red; Later-to_earlier, full step = dark red.
Figure 4. Size distribution of RATs. Boxes in the plots represent the 25-75th inner quartile range (IQR) with the median indicated by a solid black line. Lower and upper whiskers correspond to 25th quartile – 1.5*IQR and 75th quartile + 1.5*IQR, respectively.
Earlier-to-later Half step
\((n=2436)\)

Later-to-earlier Half step
\((n=1300)\)

Later-to-earlier Full step
\((n=58)\)

Earlier-to-later Full step
\((n=75)\)

H3K4me3

H3K27me3

H3K56ac

Transcription

Figure 5. Scatter plots of epigenetic and transcriptional data at RATS. The light \((x-axis)\) and dark \((y-axis)\) FE values for epigenetic marks and transcriptional data, at each 1-kb window that comprises a RAT, were plotted.
Figure 6. Distribution of epigenetic and transcriptional differences at RATs with corresponding box and whiskers plots. Boxes in the plots represent the 25-75th inner quartile range (IQR) with the median indicated by a solid black line. Lower and upper whiskers correspond to 25th quartile $-$ 1.5*IQR and 75th quartile + 1.5*IQR, respectively. Outliers are indicated by black dots that lie outside of the whiskers.
Figure 7. IGV image of Light and Dark RT segmentation with corresponding outlier RAT density track. The upper panel shows distribution of RATs across the chromosomes. The zoom displays a close view of chromosome III. Each tick mark in the outlier track represents a RAT.
Table 1. Mappable read counts from Illumina paired-end data.

<table>
<thead>
<tr>
<th></th>
<th>Biorep</th>
<th>Light</th>
<th>Dark</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RNA Seq</strong></td>
<td>1</td>
<td>2163613</td>
<td>2446007</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3388255</td>
<td>2724444</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6411363</td>
<td>1815183</td>
</tr>
<tr>
<td><strong>H3K4me3 ChIP Seq</strong></td>
<td>1</td>
<td>42647765</td>
<td>26315367</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>37944687</td>
<td>29737139</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>37315965</td>
<td>32629471</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>40730119</td>
<td>38271112</td>
</tr>
<tr>
<td></td>
<td>Input</td>
<td>29224111</td>
<td>36474278</td>
</tr>
<tr>
<td><strong>H3K27me3 ChIP Seq</strong></td>
<td>1</td>
<td>47585034</td>
<td>43642450</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>44386074</td>
<td>28444187</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>33817532</td>
<td>39049038</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>52698114</td>
<td>34743808</td>
</tr>
<tr>
<td></td>
<td>Input</td>
<td>43856509</td>
<td>35444977</td>
</tr>
<tr>
<td><strong>H3K56ac ChIP Seq</strong></td>
<td>1</td>
<td>41771588</td>
<td>20911093</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>36209233</td>
<td>35028561</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>37171210</td>
<td>42310254</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>30986616</td>
<td>31997700</td>
</tr>
<tr>
<td></td>
<td>Input</td>
<td>41976627</td>
<td>27113923</td>
</tr>
<tr>
<td><strong>Repli-Seq</strong></td>
<td>Early 1</td>
<td>6541753</td>
<td>892078</td>
</tr>
<tr>
<td></td>
<td>Early 2</td>
<td>3335710</td>
<td>2804170</td>
</tr>
<tr>
<td></td>
<td>Early 3</td>
<td>9244254</td>
<td>2878434</td>
</tr>
<tr>
<td></td>
<td>Mid 1</td>
<td>12957879</td>
<td>1934620</td>
</tr>
<tr>
<td></td>
<td>Mid 2</td>
<td>7022280</td>
<td>21402057</td>
</tr>
<tr>
<td></td>
<td>Early 3</td>
<td>7044808</td>
<td>6986435</td>
</tr>
<tr>
<td></td>
<td>Late 1</td>
<td>5256721</td>
<td>3536098</td>
</tr>
<tr>
<td></td>
<td>Late 2</td>
<td>6212639</td>
<td>6992722</td>
</tr>
<tr>
<td></td>
<td>Late 3</td>
<td>2697429</td>
<td>6392644</td>
</tr>
<tr>
<td></td>
<td>G1 Input</td>
<td>12204470</td>
<td>4985929</td>
</tr>
</tbody>
</table>
Table 2. Potential RAT classifications. “Earlier-to-later” and “Later-to-earlier” indicates the direction of RT change from light to dark. A “Half step” refers to a change from one time class to an adjacent time class when RT is considered in a linear manner (i.e. $E \leftrightarrow EM \leftrightarrow M \leftrightarrow ML \leftrightarrow L$). A “Full step” is any change between timing classes that are not adjacent.

<table>
<thead>
<tr>
<th>Direction</th>
<th>Degree</th>
<th>Example RAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earlier-to-later</td>
<td>Half step</td>
<td>Light E -&gt; Dark EM</td>
</tr>
<tr>
<td>Earlier-to-later</td>
<td>Full step</td>
<td>Light E -&gt; Dark M</td>
</tr>
<tr>
<td>Later-to-earlier</td>
<td>Half step</td>
<td>Light M -&gt; Light EM</td>
</tr>
<tr>
<td>Later-to-earlier</td>
<td>Full step</td>
<td>Light M -&gt; Light E</td>
</tr>
</tbody>
</table>
Chapter V

Future prospectus

The ultimate goal of the work presented in this dissertation was to study replication timing (RT) in plants. To this end, I profiled RT in an Arabidopsis thaliana suspension cell culture grown under light and dark conditions. In chapter II, we sought to examine the relationship between gene expression differences and changes in the levels of H3K4me3 and H3K27me3. Genic H3K4me3 and H3K27me3 marks are associated with active and repressed transcription, respectively. Most studies comparing H3K4me3 and H3K27me3 enrichment levels with gene expression data are conducted in developmentally different cell types or in a mixed population of cells under different treatments. Large changes in gene expression occur during development and are accompanied by detectable changes in epigenetic enrichment. Mixed cell populations (e.g. whole seedlings as an example) subjected to different environmental treatments also respond with large changes in gene expression that enable the adaptive response. In this case, the expression of genes involved with developmental regulation will change in response to external cues because of the overlap between developmental and environmental responsive genetic pathways. In contrast to other studies, our cells were genetically equivalent and differed only in their light and dark growth conditions. Thus, the large changes in gene expression seen in developmentally different cell lines were not expected in the light and dark grown suspension cells. Furthermore, the cells were not subjected to whole plant signaling pathways that result in major shifts in gene expression. For these reasons, we did not know if detectable differences in enrichment of H3K4me3 and H3K27me3 would accompany changes in gene expression in the light vs. dark grown cells. We found that there are differences in the levels of the two HMs at several of the most differentially expressed genes. This gave us confidence to include these genic marks for future comparison with RT data.
An unexpected aspect of data analysis is highlighted in this study. Initially we used peak calling to identify regions of the genome that maintained different levels of H3K4me3 and H3K27me3. However, results from a direct comparison of genome-wide peak calling and differential gene expression data was somewhat misleading. A gene that is differentially expressed might maintain relatively different levels of an epigenetic mark at that location. Even though the epigenetic levels are quantitatively different at that gene, a peak caller can still identify that region as enriched in both treatments. Thus, care must be taken when comparing treatment effects (e.g. light vs. dark) on different types of data (differential gene expression vs differential epigenetic enrichment) using widely available genomic tools. This is especially important given that there are spatial restrictions set by genomic coordinates (e.g. genes, transposons) that limit our analysis of overlapping comparisons. Furthermore, assignment of epigenetic peaks with the intent of implicating a regulatory relationship becomes quite tricky when genomic features lie close together (e.g. genes in Arabidopsis). Without visually inspecting the entire genome, it difficult to be sure that peaks are assigned to the feature they are regulating versus a nearby feature. Even with visual inspection this can be difficult. In conclusion, we've provided an example of why it is important to consider the use of non-standard approaches to analysis in addition to available tools when comparing different types of genomic data sets. The use of quantitative data could improve interpretation of results when comparing across data types. For example, it is possible that linear regression analysis might be more appropriate than enrichment peak count data for comparing quantitative changes in transcription against changing levels of histone enrichment across the genome.

In chapter III, a RT profile of Arabidopsis light grown cells was presented. In general, earlier replication is associated with the active marks H3K56ac and H3K4me3 and higher levels of transcription. In contrast, late replication is associated with H3K27me3 and transcriptional levels decreases as timing becomes later. We found that transcription levels were higher for early-mid (EM) than early (E). Interestingly, our
combinatorial epigenetic analysis suggested that H3K4me3 was more closely related to EM than to E. To understand how RT, transcription and H3K4me3 come together to form these results, we must consider Arabidopsis gene density. Presuming that transcription requires open chromatin and that several genes lying near one another are being transcribed, we assume that genes that are not transcribed are still in the vicinity of open chromatin. Also, keep in mind that potential origins are throughout the genome. When replication is signaled to begin, the replication machinery is recruited to chromatin. A locus that is occupied by transcriptional machinery is not likely to be accessible to the replication machinery but the chromatin surrounding the locus would be accessible. So in the moment that replication begins, we hypothesize that the most highly transcribed genes are not replicated first because of the presence of transcriptional machinery. If the replication machinery is recruited to a nearby origin and replication begins, the highly transcribed gene will be replicated in EM as the replication fork spreads from the E region. The use of live cell imaging using labeled transcriptional and replication machinery would be useful to test this hypothesis in Arabidopsis.

New findings in Arabidopsis that highlight functional links between replication, epigenetics and transcription might provide guidance for targeted study of RT (de la Paz Sanchez and Gutierrez 2009; Ding, et al. 2012; Lauberth, et al. 2013; Iglesias, et al. 2015; Del Olmo, et al. 2016). For example, a recent study provided strong evidence for the functional role of DNAP ε, the replicative DNA polymerase for leading strand synthesis, in transcriptional silencing by H3K27me3. A mechanistic link was established between the DNAP ε catalytic unit ESD7 and the PRC2 complex that regulates H3K27me3. Del Olmo et al. (2016) used H3K27me3 ChIP-PCR in esd7 mutants to show that DNAP ε affects H3K27me3 levels at the PRC-regulated FT and SOC1 loci. Physical interactions between the catalytic subunit of DNAP ε and CLF, the SET domain-containing component of PRC2, were identified by yeast two-hybrid assays and confirmed by in vitro and in vivo assays. Importantly, they showed that ESD7 binds to
the FT and SOC1 chromatin and is required for recruitment of the PRC2 complex to these loci. It was recently established that of H3K4me3 deposition requires DNAP δ, the replicative DNA polymerase for lagging strand synthesis (Iglesias, et al. 2015).

In chapter IV, we compared changes in RT profiles between light and dark grown cells. Most of the genome maintained very similar RT except for some notable differences in the pericentromeric region. Specifically, the pericentromeric region displayed evidence of increased late replication in dark grown cells relative to light grown cells. The change to a later time could be related to increased chromatin condensation. We will soon conduct an H3K9me2 ChIP-Seq experiment in the light and dark grown cells to test this hypothesis. H3K9me2 is associated with constitutive heterochromatin and it so it possible that this mark is more enriched in the pericentromere and centromere of the dark cells.

To better under RT in Arabidopsis, future studies should compare timing changes in a system that elicits major changes in gene expression. This would be most easily accomplished by comparing mixed tissue samples, such as whole seedlings, grown under different conditions. Replication timing in different cell lineages have been successful. It is possible that different cell lineages in Arabidopsis might prove useful to study replication timing. However, it is possible that even large changes in gene expression will not greatly impact RT in Arabidopsis because of its high gene density. High gene density increases the chance that transcriptional activity of one gene impacts the chromatin openness of a nearby gene that it is not transcribed. This ultimately could lead to a large euchromatic state and decrease the likelihood of heterochromatin establishment through most of the genome. This is in contrast to a genome that is less gene dense where transcription of a gene might not affect the chromatin state of its nearest neighbor because of the distance between the genes. If less transcriptional machinery occupies a space on a chromosome then heterochromatin might be more likely to form. In this case, the confounding factor of genes distance impacting
chromatin openness is removed allowing a more direct analysis of the impact of transcriptional changes on RT. To learn more about the regulation of RT in plants it might be necessary to explore other plant model systems. In an interesting twist, the characteristics that make Arabidopsis a superior model for genetics research might be the very ones that limit our ability to draw broad conclusions regarding plant DNA RT. In the future, functional studies focused on understanding the nuclear arrangement of chromatin during replication will be useful to elucidate the complex interplay between RT and related cell processes.
References


