ABSTRACT

FRANCO-RUIZ, MARIANA. DNA Replication in Plants: Origins of DNA Replication and Post-translational Histone Modifications during the DNA Synthesis phase of Arabidopsis thaliana. (Under the direction of George C. Allen and William F. Thompson).

Mapping origins of DNA replication on a genomic scale remains a major challenge in eukaryotic molecular genetics. Although a few maps exist, notably for the lower eukaryote Saccharomyces cerevisiae, it has proven difficult to map replication origins to unique loci in the complex genomes of higher eukaryotes. In large measure, this is because (in contrast to yeast and prokaryotes) higher eukaryotes seem to lack a clear consensus sequence for replication origins. However, all eukaryotes use a similar mechanism of replication, in which the pre-replicative complex binds a DNA sequence (origin), unwinds the DNA to allow the replication machinery to start producing newly synthesized DNA or nascent strands. Such similarities offer hope that genomic tools can be used to define origins of DNA replication in higher eukaryotes. The first phase of the work described here is an effort to map origins of DNA replication in Arabidopsis thaliana using three different approaches: enrichment of nascent strands labeled with bromo-deoxy-uridine (BrdU), enrichment of nascent strands by size selection and enrichment of nascent strands by the use of lambda exonuclease. DNA enriched by each procedure was mapped onto custom-designed high resolution microarrays representing Arabidopsis thaliana chromosome 4.

The nascent strands obtained with BrdU were not reproducible. For the nascent strands obtained by size selection, we got 721 putative origins and with the use of lambda exonuclease we obtained 330 putative origins. The two later methods had 217 putative origins in common. Comparison of the two dataset with features for chromosome 4 (Genes,
coding sequences, exons, transposable elements etc) showed that the two sets contain different features. Further analyses are required to identify true origins.

A second phase of the project sought to describe the abundance and distribution of various post-translational modifications of histones H3 and H4, using both Western blot and immunolocalization techniques. Modified histones play an important role in transcription and according to recent studies also in replication due to their role in chromatin structure. We found out that histone 4 acetylated in the lysine 5 and 16, and histone 3 acetylated in the lysine 56 do not present changes during the cell cycle in our cell culture. We also observe that histone 3 dimethylated in lysine 9 and phosphorylated in serine 10, present very low abundance and showed signal in the nucleolus. The data provide a basis for a future analysis of how these epigenetic marks are transmitted during DNA replication and mitosis.
DNA Replication in Plants Origins of DNA Replication and Post-translational Histone Modifications During the DNA Synthesis Phase of Arabidopsis thaliana

by
Mariana Franco Ruiz

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

Horticultural Science

Raleigh, North Carolina

2012

APPROVED BY:

Dr. George C. Allen 
Chair of Advisory Committee

Dr. William F. Thompson 
Vice-Chair of Advisory Committee

Dr. Linda Hanley-Bowdoin

Dr. Bryon R. Sosinski
DEDICATION

“La mejor herencia que les puedo dar es la educacion” (“The best legacy I can give you is your education”) - Jorge A. Franco-Lopez

To Trino and Luka, my best decisions

To my grandparents, who were always interested in learning more.

To my Parents, Sofia and Jorge, your care, your understanding, your support and your love make me.

To my Siblings: Martalicia, Jorge and Mosy, whose love and teaching always guided me.

To my Mother in Law, Elena, who believes in me

To my Raleigh Family: Dolores and Bill, who understood my need for a family in a foreign country. Miguel and Yokiko, for being great friends and for loving and caring for my son.
BIOGRAPHY

Mariana Franco Ruiz was born and raised in Salamanca, Guanajuato, México, on December 3rd, 1975 to Jorge A. Franco-Lopez and Sofia V. Ruiz-Rojas. Mariana has 3 older siblings, Martalicia, Jorge and Montserrat. Mariana obtained her Engineering degree in Food Science at the ICA (Institute for Agricultural Sciences, a Universidad de Guanajuato campus) in Irapuato, México in 1999. From her days in high school Mariana has been interested in getting an education that can help in solving the food problems around the world. Mariana’s bachelor’s degree thesis work, done in CINVESTAV, Irapuato, put her in touch with researchers, whom were solving agricultural problems, which encourage her to join the scientific community.

In 2001 Mariana came to Raleigh NC, USA, married to J.T. Ascencio-Ibáñez who was attending his PhD at North Carolina State University. At NCSU, Mariana met Dr. Niki Robertson. Dr. Niki Robertson opened the doors for Mariana and helped her establish a connection with Dr. George C. Allen. Dr. Allen hired Mariana as a research scholar from 2002 to 2005. At the end of 2005 she decided to pursue her PhD, under the guidance of Dr. George C. Allen and Dr. Bill Thompson, in the very attractive project on “DNA Replication”. Thanks to DNA replication, Mariana had a wonderful son, Luka Franco Ascencio, born in October 2008.
ACKNOWLEDGMENTS

First I want to thank Dr. Dominique Robertson for being the key that opened the doors for me at North Carolina State University. I also want to thank you for your support and because it means a lot to be appreciated by a great person, thank you so much Niki!

I also want to thank Drs. George Allen and Bill Thompson for all their encouragement and the time they invested in me during my research.

To all my committee members, Dr. Linda Hanley-Bowdoin, Dr. Bryon Sosinski, Dr. Bill Thompson and Dr. George Allen, I want to thank you because you were always there for me. You guided me and helped me to grow as a professional. You are a fantastic team and there is so much to learn from all of you.

A very special thank you to Dr. Pete Pascuzzi, for your incredible work analyzing my data and for those numerous discussions in science.

I want to thank the lab, the Plant Gene expression lab. Tae-Jin Lee, Miguel A. Flores-Vergara, Rosy Hatch, Emily Wear, Yokiko Hiromoto, A Chin, Shingo Nagaya, and from the Plant Transformation lab, Sergei Krasnyanski. Guys thank you very much for your friendship and for your help during my research.

Thank you to the Viral Genomics lab, Mary Beth Dallas, Maria Reyes, Leandro de Leon-Guerra, Ashley Brooks, Pete Pascuzzi, Emily Markham, Laura Greeley and Wei Shen, for sharing lab space with me and giving me ideas to solve technical problems.
Thank you Rachel McLaughlin for all your time answering my questions, solving problems, listening to me, doing things that are not part of your job… I can go on and on how important you are… you made me feel at home in the Horticultural Science Department.

Thank you Dr. John Dole, for all your support during my research and for all the advice and for allowing me to improve your spanish, I had fun.

To Dr. Julia Kornegay, Thank you for all you support as head of department and as director of graduate programs.

Thank you so much to Dr. Eva Johannes, for your invaluable time in the confocal microscope and your patience to teach me.

Thank you Dr. Linda Hanley-Bowdoin for your kind support, your time and for sharing your lab space with me.

Thank you OISS staff, your help and your friendship is essential for foreigners at NCSU, especially thank you Hanya!

Thank you Dr. Tara Nash, you are an incredible friend and thank you for your shoulder while I was finishing this dissertation, you are great.

Thank you Maria Reyes, the support you gave me and your friendship in the lab was invaluable for me.

Thank you very, very much Dr. Trino Ascencio for being my personal consultant during 9 years and for always believing in me.

Last but not least, I thank NSF and the Graduate School for the financial support.
# TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................... ix
LIST OF FIGURES .................................................................................................................... x

CHAPTER I Eukaryotic DNA Replication: Characterization of Origins and Histone
Posttranslational Modifications ............................................................................................... 1

INTRODUCTION ....................................................................................................................... 2
- The Fundamentals of Origins of DNA Replication: From Prokaryotes to Eukaryotes ........ 2
- DNA Replication in *S. cerevisiae* ................................................................. 3
- DNA Replication in Metazoans ................................................................................... 6
- Plants .............................................................................................................................. 8
- What Constitutes an Origin? ....................................................................................... 10
- Initiation Zones vs Discrete Origin ......................................................................... 11
- Origin Usage and Efficiency ..................................................................................... 11
- Mapping Origins ......................................................................................................... 13
- Nascent Strands ........................................................................................................... 13
  - BND Cellulose ....................................................................................................... 14
  - Bromo-deoxy-uridine ............................................................................................ 14
  - Lambda Exonuclease .............................................................................................. 14
- Two-Dimensional (2-D) Gels .................................................................................... 15
- Bubble Trapping and Matrix Attachment Regions ..................................................... 16
- RIP (Replication Initiation Point) ................................................................................ 19
  - In situ Approaches .................................................................................................. 21
    - Fiber Autoradiography ......................................................................................... 21
    - Chromatin Immunoprecipitation (ChIP) ............................................................. 22
    - Origin Analysis ..................................................................................................... 24
    - Episomes ............................................................................................................... 25
- Differences Between Plants and other Higher Eukaryotes ........................................... 26
- Validation of Putative Origins ..................................................................................... 28
- Origins of DNA Replication and Chromatin Structure ............................................. 28
- Replication of Centromeres and Pericentromeric Heterochromatin ......................... 32
- Research Rationale .................................................................................................... 34
- REFERENCES .................................................................................................................. 36
- FIGURES .......................................................................................................................... 54

CHAPTER II. Mapping Origins of DNA Replication from the Distal Long Arm of
*Arabidopsis thaliana* Chromosome 4 .............................................................................. 61

INTRODUCTION ....................................................................................................................... 63
MATERIAL, METHODS AND NOTES ..................................................................................... 68
CHAPTER III. Changes in Histone Modification During G1, S and G2/early M .......126
INTRODUCTION .....................................................................................127
MATERIALS, METHODS AND NOTES ....................................................134
  Cell Culture .......................................................................................134
  Nuclei Isolation .................................................................................134
  Flow Sorting and Protein Extraction ..................................................135
  Western Blots and Statistical Analysis ................................................135
  Cell Fixation for Immunolocalization of Histone Modifications ............136
  Nuclei Isolation for Immunolocalization of Histone Modifications ........137
  Nuclei Sorting for Immunolocalization of Histone Modifications ..........137
  Immunolocalization of Histone Modifications ......................................137
  Confocal Microscopy .........................................................................138
  Notes ...............................................................................................139
RESULTS .............................................................................................140
Quantifying Modified Histones in G1, S and G2 Nuclei ........................................ 140
Localization of the Posttranslational Marks During Interphase .......................... 141
DISCUSSIONS ........................................................................................................ 144
H4K5ac, H4K16ac, H3K56ac, and H3S10p Levels Remain Unchanged During the
Cell Cycle ........................................................................................................ 144
Subnuclear Localization of Histone Modifications During the Cell Cycle .......... 146
H3K27me3 and H3K9me2 Variation During the Cell Cycle ............................ 147
REFERENCES .................................................................................................... 150
FIGURES AND TABLES .................................................................................. 158

CHAPTER IV. Prospectus ..................................................................................... 169
PROSPECTUS .................................................................................................... 170
The Dynamic Epigenome .................................................................................. 174
REFERENCES .................................................................................................... 176
LIST OF TABLES

CHAPTER II:
Table 1. Genomic features and PTMs for Neat and Exo SNS.........................................................99

CHAPTER III:
Table 1. Antibodies use..................................................................................................................158
Table 2. Paired T-test analysis of modified histones in G1, S and G2/early M nuclei....159
Table 3. Antibody dilutions ...........................................................................................................160
LIST OF FIGURES

CHAPTER I:
Figure 1. Origin of DNA replication and nascent strands ..............................................54
Figure 2. Two dimensional (2-D) gels ...........................................................................55
Figure 3. Matrix Attachment Regions (MARs) .................................................................56
Figure 4. Replication Initiation Point (RIP) mapping .........................................................57
Figure 5. Single Molecule Analysis of Replicated DNA (SMARD) .....................................58
Figure 6. Chromatin Immunoprecipitation (ChIP) .............................................................59
Figure 7. Chromatin structure .........................................................................................60

CHAPTER II:
Figure 1. Flow diagram BrdU SNS .................................................................................100
Figure 2. Flow diagram Neat SNS ..................................................................................101
Figure 3. Flow diagram Exo SNS ....................................................................................102
Figure 4. BrdU SNS correlation between microarray bioreplicates .................................103
Figure 5. Neat SNS correlations ....................................................................................104
Figure 6. Exo SNS correlations ......................................................................................105
Figure 7. Comparison of SNS peaks ..............................................................................106
Figure 8. Neat and Exo SNS comparison ......................................................................107
Figure 9. Venn diagram: Agreement between Neat and Exo SNS .................................108
Figure 10. Percentage overlap between Exo and Neat SNS peaks .................................109
Figure 11. AT content for Neat, Exo and overlapping SNS ............................................110
Figure 12. Unannotated features for Exo, Neat and overlapping SNS .............................111
Figure 13. Transcription Start Site (TSS) and Transcription Termination Site (TTS) for Neat and Exo SNS ..........................................................112
Figure 14. SNS position relative to proximal gene ...........................................................113

CHAPTER III:
Figure 1. Immunoblot for modified histones ................................................................161
Figure 2. Confocal microscopy images of sorted Arabidopsis nuclei in G1, S and G2/early M phase .................................................................162
CHAPTER I

Eukaryotic DNA Replication: Characterization of Origins and Histone Posttranslational Modifications
INTRODUCTION

The Fundamentals of Origins of DNA Replication: From Prokaryotes to Eukaryotes

Much of our current understanding of DNA replication initiation is based on the “replicon model” originally proposed by Jacob and Brenner (Jacob & Brenner, 1963). The replicon model postulates that a trans-acting factor (initiator) recognizes a specific genetic element (replicator) to initiate DNA replication of a defined unit (replicon) of the genome. The model proposes the existence of a regulator (protein) for DNA replication that recognizes a specific site (origin) to start replication. The replicon model was validated using *Escherichia coli* (*E. coli*) to identify the protein (DnaA) that melts and stabilizes DNA to initiate replication, DnaC as the helicase regulator, DnaB as the replicative helicase protein and oriC as the origin of DNA replication. OriC is ~250 bp long, AT-rich, and includes three to four repeats of a consensus sequence of 9 bp 5’ TTATC/ACAC/AA, called the DnaA box (Robinson & Bell, 2005). Several monomers of DnaA protein bind to the DnaA box, which then melts and stabilize the DNA to form the replication bubble (Robinson & Bell, 2005). Upon formation of the replication bubble, DNA polymerases start replicating the genome – a complex and fascinating process that is outside the scope of this review. While the simple prokaryotic model can also be applied to many viruses and the simple eukaryotic model system, *Saccharomyces cerevisiae* (*S. cerevisiae*), it is not fully applicable to higher eukaryotes (Figure 1).
DNA Replication in *S. cerevisiae*

The first eukaryotic origin was found in the budding yeast *S. cerevisiae* (Stinchcomb, Struhl, & Davis, 1979). A putative replicator sequence (origin) was inserted into a bacterial plasmid, which was then used to transform yeast. If the sequence being tested supported autonomous plasmid replication, it was considered to be an autonomously replicating sequence (ARS) and a putative origin, which led to the development of an assay for routine mapping of potential origins of DNA replication. Although the ARS plasmid assay does not measure origin activity in a chromosome, subsequent 2-D gel analyses showed that several ARSs could also function as chromosomal origins of DNA replication (Brewer & Fangman, 1987; Huberman, Spotila, Nawotka, el-Assouli, & Davis, 1987). Such experiments helped to define yeast origins as chromosomal loci, but many of the key proteins required for DNA replication remained unknown. In a ground-breaking study it was identified the elusive eukaryotic initiator protein complex (ORC, origin recognition complex) using a combination of DNase foot printing and DNA affinity chromatography (S. P. Bell & Stillman, 1992). The same group also identified the six distinct subunits of the origin recognition complex (ORC1-6). Mutations in the four key sequence elements of ARS-1 (A, B1, B2 and B3) showed that ORC binds to the ARS “A” element and proved that ORC is required for DNA replication (S. P. Bell & Stillman, 1992).

The chromosomal origins of DNA replication of *S. cerevisiae* range from 100-200 bp in length, and contain three or four sequence elements (A, B1, B2 and B3) associated with ARS1 origin function. Elements B1, B2, and A are essential for DNA replication initiation in the chromosome (Newlon & Theis, 1993). Element A contains the ARS consensus
sequence, or ACS, (5’-A/TTTTAT/CA/GTTTA/T-3’), which constitutes half of the site where ORC binds. The other half of the ORC binding site is contained in the B1 element (Rowley, Cocker, Harwood, & Diffley, 1995). Some *S. cerevisiae* origins, such as ARS101 and ARS310, are compound origins and contain several ACSs. The presence of any one of the ACSs is adequate for replication of the region (Cvetic & Walter, 2005). Interestingly, the organization of the *S. cerevisiae* compound origins is more similar to the origins of higher eukaryotes, where a region of DNA (“initiation zone”) may contain several origins, any one of which may fire to start replication within the zone. While *S. cerevisiae* has compound origins, the origins are still simpler to map than the origins of higher eukaryotes because all *S. cerevisiae* origins have the well-defined ACS (Cvetic & Walter, 2005).

New technologies for genome-wide analysis, along with a comprehensive knowledge of the *S. cerevisiae* replication machinery, have made it possible to map origins throughout the *S. cerevisiae* genome (Wyrick et al., 2001; Xu, Aparicio, Aparicio, & Tavare, 2006). Initial studies to identify origins of replication in the *S. cerevisiae* genome used polyclonal antibodies to ORC (1 to 6), along with minichromosome maintenance proteins (Mcm 3, 4, and 7) for Chromatin Immunoprecipitation (ChIP) and subsequent co-mapping on microarrays to identify the replication origins within the entire *S. cerevisiae* genome (Wyrick et al., 2001). While ChIP against ORC and MCM worked well for mapping origins in *S. cerevisiae* due to the ACS, it was found that the Drosophila ORC2 localizes to specific chromosomal locations, but no consensus sequence could be identified (MacAlpine, Gordan, Powell, Hartemink, & MacAlpine, 2010).
The origins of DNA replication in *Schizosaccharomyces pombe* (fission yeast) are more complex than those of *S. cerevisiae*. Early studies (Beach & Nurse, 1981) tested several DNA sequences from three different microorganisms for their ability to support autonomous replication – and hence transformation - of *S. pombe*, and found that the *S. cerevisiae* ARS could not act as origin in *S. pombe*. Perplexingly, two plasmids containing bacterial and parts of the 2µm plasmid of *S. cerevisiae* but lacking chromosomal DNA from *S. pombe*, were able to yield high frequency transformation in *S. pombe*. It is important to note that none of the *S. pombe* chromosomal DNA fragments used in this study were tested as origins of DNA replication in a chromosomal context. Subsequent studies showed that some chromosomal sequences in *S. pombe* could promote autonomous replication in the same organism, but the *S. pombe* origins defined in this way did not show a clear consensus sequence such as that in *S. cerevisiae* (Losson & Lacroute, 1983; Toda, Nakaseko, Niwa, & Yanagida, 1984). The same technique used by Beach and Nurse in 1981 was then used to map ARS sequences in fission yeast (Dubey, Zhu, Carlson, Sharma, & Huberman, 1994; Johnston & Barker, 1987; Maundrell, Hutchison, & Shall, 1988).

The fine structure of the *S. pombe* ARS was first described in 1995, it was also called ARS1 as in *S. cerevisiae*, because it was also the first origin identified in *S. pombe* (Clyne & Kelly, 1995). The *S. pombe* ARS1 is approximately 0.5 to 1 Kb and contains a 20-50 bp block of asymmetric A-T stretches that are important for ORC binding and DNA unwinding. It is interesting that ARS sequences from *S. cerevisiae* do not function as replication origins in *S. pombe* and vice versa (T. J. Kelly & Brown, 2000). Today it is known that there is not a consensus sequence to distinguish *S. pombe* origins of DNA replication. Dai et al. (Dai,
Chuang, & Kelly, 2005) proposed that replication in *S. pombe* initiates stochastically from A-T stretches at intergenic regions, which may also serve as a model for metazoans.

**DNA Replication in Metazoans**

An origin of DNA replication in the dihydrofolate reductase (DHFR) region in a Chinese hamster ovary (CHO) cell line was uncovered in 1982 (Heintz & Hamlin, 1982). The DHFR origin has been used as a model because it is amplified 500 times, which allows efficient incorporation of $^{14}$C-thymidine in synchronized cells. When the $^{14}$C-thymidine-labeled genomic DNA was digested and the fragments analyzed by autoradiography, a specific initiation site was identified (Heintz & Hamlin, 1982). However, more recent analyses of the DHFR region (Anachkova & Hamlin, 1989; Kobayashi, Rein, & DePamphilis, 1998; Pelizon, Diviacco, Falaschi, & Giacca, 1996) indicated that the specific initiation site is more complex. A variety of techniques were used and consistently showed that the initiation site described by Heintz, et al (1982) is actually an initiation region containing several different replication origins in close proximity. To detect origin-enriched regions, the DNA of the CHO cells was arrested in G1 phase, UV-crosslinked and labeled with $^3$H-thymidine and the DNA was isolated, fractionated to collect for the low molecular weight DNA that was analyzed on a dot blot probed with restricted fragments from the DHFR region (Anachkova & Hamlin, 1989).

Pelizon et al, (1996) used short nascent strands (SNS) and bromo-deoxy-uridine (BrdU) incorporation to enrich for sequences with putative origins, which was then confirmed with competitive PCR. A modified form of the SNS approach uses λ exonuclease
treatment to digest the DNA that is not protected by an RNA-primer at the 5’end. The resulting nascent strands were then analyzed with competitive PCR for the presence of origins (Kobayashi et al., 1998). Finally, data from 2-D gel analysis led to the conclusion that the DHFR region was a replication zone with multiple origins (Dijkwel, Mesner, Levenson, d'Anna, & Hamlin, 2000; Dijkwel, Vaughn, & Hamlin, 1994; Vaughn, Dijkwel, & Hamlin, 1990; Wang, Dijkwel, & Hamlin, 1998).

Additional studies have found origins of replication using different approaches, organisms and cell lines. Examples of such origins of replication include the c-myc locus from HeLa cells (McWhinney & Leffak, 1990), the Lamin B2 locus from human cells (Biamonti et al., 1992), the β-globin locus (Kitsberg, Selig, Keshet, & Cedar, 1993), and the Igh locus from mouse (Ariizumi, Wang, & Tucker, 1993). Particularly noteworthy are studies of early embryos of Drosophila (Shinomiya & Ina, 1991) and Xenopus (Hyrien & Mechali, 1993), where it was shown that DNA replication starts at random sequences of the chromosome at regular intervals. However, it was also found that when Xenopus was tested at a later developmental stage, the origins are defined (Bozzi, Baldari, Amaldi, & Buongiorno-Nardelli, 1981). This result suggests that origins change positions according to developmental states, which may be due to changes in transcriptional activity. Early Xenopus embryos are transcriptionally inactive and the cell cycle is very rapid, but upon entering the midblastula stage transcription resumes and the cell cycle slows (Hyrien & Mechali, 1993).

A combination of nascent strand abundance and a tiled microarray was used to identify 28 new origins of replication from three 50-Kb regions in asynchronous human cells,
in addition to confirming four that had been previously identified (Lucas et al., 2007). The use of genomic tools now makes it possible to map potential origins at a global scale.

However, despite the development of new technologies, the locations of origins in higher eukaryotes remain uncharacterized due to a lack of a consensus sequence and the regulation of DNA replication is still not understood. Without the knowledge of a specific sequence, assuming that a specific sequence even exists, finding origins in metazoans will remain difficult with only a handful that have been characterized.

**Plants**

DNA replication in plants is even more poorly understood. Early studies by Van't Hof, et al (1987) described an origin (ori-r9) located within the ribosomal DNA (rDNA) region of *Pisum sativum*. Ori-r9 which has a proximally located ARS consensus sequence (Hernandez, Lamm, Bjerknes, & Hof, 1988), was confirmed to be an origin by 2-D gels analysis (Van't Hof & Lamm, 1992). Since the original discovery of ori-r9, many unsuccessful attempts have been made to map origins of DNA replication in plants (J. Bryant, personal communication).

Murata, et al (2005) used fluorescence microscopy to identify newly replicated DNA from carrot somatic embryos BrdU-labeled. The BrdU-labeled regions were then isolated using laser microdissection for subsequent cloning into a “DNA replication origin” micro-library. To our knowledge, none of putative replication origins isolated by Murata, et al (2005) study have been confirmed by other techniques. The use of laser microdissection in an unsequenced genome, for isolation of newly replicated BrdU-labeled DNA appears to be
tedious and lacks the necessary sensitivity and resolution to identify actual origins of DNA replication. A more feasible approach is the isolation of BrdU-labeled DNA by immunoprecipitation, as described later in this thesis.

Although there is limited information on replicons of different plant species, the knowledge that plant S phase can be shortened by different factors like floral stimulus (Durdan SF, 1998; Jacqmard & Houssa, 1988), hormones such as cytokinin (Houssa, Jacqmard, & Bernier, 1990) and developmental stages related to transcription (Bryant, 2010) suggests that additional origins can be activated under these conditions. Mapping origins in plants will be more difficult if origin usage varies extensively (for an excellent review see (Bryant & Aves, 2011). In addition, a recent genome-wide bioinformatic study of the replication machinery in Arabidopsis and rice showed that the proteins involved in plant DNA replication are more similar to vertebrates than they are to budding yeast (Shultz, Tatineni, Hanley-Bowdoin, & Thompson, 2007). Plants have some unique characteristics, including several altered core replication genes, multiple copies of the RPA genes, duplicate copies of ORC1, CDC6 and CDT1, and other genes required for elongation of replication (Shultz et al., 2007). Antibodies for ChIP on chip of the plant pre-replicative complex are not commercially available. As a consequence, antibodies for plants studies must be generated and validated for qualitative analyses.

Importantly, models of DNA replication in plants should also consider that the plant cell cycle includes endoreduplication, in which the genome is duplicated without passing through mitosis. Endoreduplication is a common process in eukaryotes, that occurs widely in many cell types of mammals and arthropods, as well as in plants. Endoreduplication has been
well studied in maize endosperm (Lopes & Larkins, 1993; Lur & Setter, 1993) and is known to occur in every cell type of dicotyledonous plants with the exception of gametes, guard cells and meristems (De Veylder, Beeckman, & Inze, 2007; Melaragno, Mehrotra, & Coleman).

**What Constitutes an Origin?**

In bacteria, an origin defines the DNA sequence where DNA replication starts for subsequent duplication of the entire circular chromosome. However, in eukaryotes an origin initiates replication in only a fraction of the linear chromosome, and complete replication of the entire genome requires multiple origins. The bacterial origin, (oriC), contains two main elements, which include a set of tandem repeats where DnaA protein binds to alter the DNA conformation, and an AT-rich region that allows the separation of the strands to facilitate binding of the helicases and polymerases that start the actual replication process (Robinson & Bell, 2005).

ARS activity in budding yeast requires the presence of at least three of four functional conserved elements, which include A, B1, B2 and the enhancer B3 element. The A and B1 elements are needed for ORC binding, along with a DNA unwinding element (DUE) that is located in the B2 element. Some ARS have other elements such as binding sites for transcription factor Abf-1 (DePamphilis, 1999). Abf-1 is a DNA binding protein required for transcriptional activation, DNA replication, DNA repair and gene silencing (Ganapathi et al., 2011).
**Initiation Zones vs Discrete Origins**

Metazoans primarily use initiation zones rather than discrete origins (Aladjem, 2007). Zones of initiation are regions of varying size (i.e. 55 Kb for CHO DHFR, 31 Kb for human rDNA and 5 Kb for Drosophila histone gene) containing multiple sites at which DNA replication can initiate in any given round of replication. Conversely, discrete origins are specific DNA sequences where replication starts at the same point in each cycle. The preferred sites used in initiation zones may be controlled by a variety of factors such as developmental or environmental cues, and epigenetic modifications. Some origins tend to cluster in zones, such as c-myc in human (Waltz, Trivedi, & Leffak, 1996) or Xic in mouse (Rowntree & Lee, 2006). To differentiate discrete sites from true zones, Aladjem et al (2007) proposed that zones include origins of similar efficiency whereas discrete origins are start sites from where initiation occurs at a high frequency (Aladjem, 2007). An alternative view is that there are preferred DNA synthesis starts sites within initiation zones (Gerbi, S. personal communication). A recent review suggested that the mammalian genome has many degenerate and redundant origins with some having evolved to become efficient and discrete (Hamlin, Mesner, & Dijkwel, 2010).

**Origin Usage and Efficiency**

It is known that only a fraction of potential origins fire during each cell cycle (Gilbert, 2010), which suggests that the firing of some origins is preferred over others. This pattern is not fixed and changes during development or physiological stress. Efficiency indicates the probability of a potential origin to initiate replication, while usage is which origin site is used
at a specific time during development. The mouse immunoglobulin heavy chain locus region ($Igh$), represents an example of reprogrammable replication timing (Norio et al., 2005). When the $Igh$ locus is inactive, DNA replication occurs as triphasic timing - i.e., some cells in the population replicate this locus early in S phase, others in mid-S and still others in late S phase. However, during B cell development, the $Igh$ locus is activated and replicates during early S phase in all cells. The use of a modified Fluorescent In Situ Hybridization (FISH) and SMARD (Single molecule analysis of replicated DNA) analysis suggested that the change in replication timing at this locus is due to the activation of origins and not to an increase or decrease in replication fork speed (Norio et al., 2005).

As an example of usage, it was found that when the three preferred origins (ori-ß, ß’ and γ) were deleted from the $DHFR$ locus, the replication timing of the region did not change (Mesner, Hamlin, & Dijkwel, 2003). They suggest that alternate origins that normally do not fire are activated to replicate the region when the preferred origins are deleted (Mesner, Hamlin, & Dijkwel, 2003). The concept of alternate origin usage is also supported by the results of a study with mouse embryonic stem cells, in which the deletion of a sequence with CpG islands containing a known origin, resulted in a change in activation of unused origins (Rowntree & Lee, 2006). In addition, Anglana, et al (2003) found that altered nucleotide availability causes a change in origin usage, as decreases in nucleotide pool levels resulted in the activation of normally dormant, alternate origins in CHO cells.
Mapping Origins

Several techniques to map origins have been developed over the years. Some are based on the structure of the replication process such as bubble trapping, or the DNA binding proteins like ChIP. Some other techniques are based on the characteristics of the nascent strands like Lambda-exonuclease, or sequence position within a MAR-bound loop, or the incorporation of new nucleotides etc. All these techniques have been used to map origins, although none has been shown to work for all organisms.

Nascent Strands (NS)

After replication is initiated, two different types of short daughter molecules are immediately produced. One type is the Okazaki fragments, which are synthesized discontinuously from the lagging strand, and range from approximately 100 to 2000 nt in length. These are then joined by DNA ligase 1 to produce a continuous lagging strand. The second, termed short nascent strands are the products of leading strand synthesis, which is a continuous process leading directly to a large daughter molecule (Figure 1). For mapping origins, one can exploit the fact that these nascent strands go through a phase in which they are longer than Okazaki fragments but short enough to be close to the origin from which they were initiated. Thus, if one can isolate short nascent strands slightly longer than Okazaki fragments and map them to the genome using microarrays or deep sequencing techniques, it should be possible to locate the positions of origins throughout the genome. There are several methods that have been used to enrich for short nascent strands, which I will review below.
**BND Cellulose** (benzoylated naphthoylated DEAE-cellulose) has been used to bind single stranded DNA (ssDNA), RNA and DNA containing single stranded regions, which allows for the separation of nascent strands from completely double stranded DNA (dsDNA). Size selection can then be used to remove the Okazaki fragments, which are believed to range from 100 to 300 nt in mammals and from 1 to 2 Kb in bacteria (Georgescu et al., 2009; Heller & Marians, 2005; Keller, Ladenburger, Kremer, & Knippers, 2002; McInerney & O'Donnell, 2007; Yao, Georgescu, Finkelstein, & O'Donnell, 2009). The Okazaki fragments size in plants is not known. The first and only plant origin described thus far was found using BND-cellulose-enriched nascent strands followed by DNA gel blot analysis (Van't Hof, Hernández, Bjerknes, & Lamm, 1987).

**Bromo-deoxy-uridine** (BrdU, a thymidine analogue) incorporation has been used to enrich for short nascent strands by selecting only very recently synthesized DNA. BrdU is added to either synchronous or asynchronous cells for a defined, very short period of time (pulse). It then becomes incorporated into the newly synthesized DNA at various thymidine positions. The BrdU-labeled DNA is fractionated by size using an alkaline sucrose gradient centrifugation and each fraction is measured for DNA size using an alkaline gel that keeps the DNA single-stranded. The preferred size range, which is greater than the size of the Okazaki fragments, is then immunoprecipitated using anti-BrdU. The nascent strands obtained by this method can be analyzed by competitive PCR (Falaschi & Giacca, 1994), tiling microarrays (Lee et al., 2010) or high throughput sequencing.

**Lambda Exonuclease.** The lambda exonuclease procedure is based on the fact that leading strand synthesis is initiated from an RNA primer. Nascent leading strands have a
short stretch of RNA located at their 5’ ends. Lambda-exonuclease digests DNA strands from 5’ to 3’, Newly synthesized DNA strand containing the 5’ RNA primer sequence is protected from digestion. Thus, you can enrich for nascent strands with lambda exo nuclease and analyze them for origins either by hybridizing them to a tiling microarray or sequencing the enriched nascent strands.

Lambda exonuclease based enrichment has been used for replication initiation point (RIP) mapping (Gerbi & Bielinsky, 1997), which is described below. Importantly, the Lambda exonuclease procedure also requires purification of the nascent strands away from the Okazaki fragments, because the Okazaki fragments are also protected from the lambda exonuclease by an RNA primer sequence. Okazaki fragments generally range in size from 100 to 200 bp in different eukaryotic organisms (Giacca et al., 1994). However, because the actual size of Okazaki fragments in plants is unknown, we used nascent strands above 1 Kb for mapping to ensure that nascent strands were separated from the Okazaki fragments. In addition, the isolated nascent strands were kept to less than 3 Kb in order to be in close proximity to the replication origin.

**Two-Dimensional (2-D) Gels**

The 2-D gel is generally considered to be the “gold standard” for confirming origins of DNA replication. However, the use of 2-D gels is limited to regions where a putative origin is likely to be present. Two main techniques are used for 2-D gel analysis of origins. The first was developed by Huberman’s lab (Huberman et al., 1987) and the second was developed by Fangman’s lab (Brewer & Fangman, 1987). Both techniques use restriction
enzyme-digested genomic DNA and separated by mass using conventional agarose gel electrophoresis (first dimension). The lane is excised, and inserted at the top of a second agarose gel that has been rotated by 90° to separate DNA fragments based on structure (second dimension). The conformation of the DNA during replication forms a Y or an O structure (like the replication bubble) that migrate at different speeds through the gel creating different patterns. The 2-D gel is then analyzed by DNA hybridization using a probe for the region containing the putative origin. Huberman’s technique differs from Fangman’s in that it can detect the direction of the fork can be determined by using short sequences to probe. The use of 2-D gels is technically challenging, and considered to be a “state of the art assay” (Hamlin, personal communication) as only a few labs are presently able to get results that can be clearly interpreted. Examples of patterns diagnostic for the presence of a replication fork are shown in Figure 2. Repeats, large amounts of genomic DNA, and random single strand nicks can affect the results, and enrichment for nascent strands before performing 2-D gels can result in better resolution.

**Bubble Trapping and Matrix Attachment Regions**

Matrix attachment regions (MARs) provide a highly organized nuclear framework (Figure 3), which facilitates higher-order chromatin structure to coordinate critical nuclear processes such as replication and transcription (Allen, Spiker, & Thompson, 2000). The nuclear matrix, or scaffold, is composed of proteins and RNAs (Fey, Krochmalnic, & Penman, 1986), although debate remains about the identity of the major components of the matrix (Mika & Rost, 2005). There is even debate about the existence of a nuclear matrix as the isolation procedure may cause artifacts, but an increasing amount of evidence suggests
that it does exist [reviewed in (Anachkova, Djeliova, & Russev, 2005)](Byrd & Corces, 2003; Fraser & Bickmore, 2007; Hancock, 2000; Nickerson, 2001).

In addition to serving a role in chromatin organization, MARs are thought to have other roles. Dijkwel et al, (1988) showed that MARs are within the DHFR locus with one located between ori-β and ori-β’, which are proposed to regulate replication. However, Mesner et al (2003) later demonstrated that deletion of the ori-β and ori-β’ MAR did not affect origin activity. Other groups have shown that MARs play a different role in replication. Djeliova et al (2001) suggested that the origins of DNA replication are attached to the nuclear matrix at some point during late G1 phase and therefore the nuclear matrix plays a role to both organize chromatin and to regulate replication activity. Recently, data was presented that corroborate the earlier work by showing that the origin attaches to the matrix during mitosis and G1 and is released from the matrix as replication proceeds (Courbet et al., 2008).

The procedure for isolating MARS consists of isolating nuclei, removing the nuclear envelope and histones to form a loose structure (called a nuclear halo) in which the now maximally extended DNA forms many loops radiating out from attachment points on the nuclear matrix. DNA in the haloes can then be cut with restriction enzymes and the fraction not bound to the nuclear matrix can be removed by centrifugation and washing. The matrix proteins are then proteinase-digested, leaving the intact MAR fragments for subsequent analysis (Figure 3). To determine if MARs are associated with origins of DNA replication the MAR isolation should be done just before or during, active replication. While a subset of MARs appear to act as origins of replication in animal cells, it is not known if the model
proposed by Courbet et al, (2008) will also apply to plants. A recent study (Flores-Vergara, 2012) that analyzed MARs in Arabidopsis cell suspension cultures, found that MARs are diverse in both sequence and function. According to this study, MARS consist in general of two populations, one rich in AT and another population rich in GC that are primarily associated with transposable elements (TEs) and with exons, respectively. Some MARs were also associated with the posttranslational mark H3K56ac, and this mark has been suggested to be associated with early replication in Arabidopsis (Lee et al., 2010).

Hamlin’s group recently reported an origin isolation technique in which matrix associated DNA is first separated from total DNA, and a bubble-trapping technique (described later) is then used to trap fragments containing replication bubbles (Mesner, Crawford, & Hamlin, 2006) (Mesner & Hamlin, 2009). Bubble trapping relies on two major purification steps, which include the isolation of supercoiled fragments that are attached to the nuclear matrix (MARs) (Pemov, Bavykin, & Hamlin, 1998) and benzoylated napthoylated DEAE (BND)-cellulose that selectively binds DNA fragments, such as replication forks containing single-stranded regions. By combining the MAR-enrichment, the BND-cellulose and bubble trapping steps, Mesner and Hamlin (2009) found a 200-fold enrichment for origins of replication, which was confirmed by 2-D gel analysis, as described earlier.

Bubble trapping is based on the property that restriction fragments containing replication bubbles can be rendered electrophoretically immobile (trapped) when mixed with molten agarose that is then allowed to polymerize. At the molecular level it is thought that strands of agarose extend through the replication bubble and immobilize the DNA fragment,
which can then be separated from linear fragments using agarose gel electrophoresis. Previously enriched fragments from the matrix binding and BND cellulose purification steps (Mesner, Dijkwel, & Hamlin, 2009) are first mixed in liquid (pre-gelled) low-melting agarose. The agarose is allowed to gel and the circular and linear fragments are separated by electrophoresis (Mesner & Hamlin, 2009). The purified circular fragments are isolated from the agarose gel and used to construct a plasmid library of putative origins. A major advantage of bubble trapping is that the bubbles can be cloned and sequenced without the need for a sequenced genome, thereby allowing for amplification and a renewable resource (Hamlin, personal communication).

Although the bubble-trapping method does not inherently require the use of the 2-D gels, Mesner, et al used the 2-D gel analysis to confirm putative replication origins and to determine whether they are initiation zones or discrete origins (for an excellent review see (Hamlin et al., 2008)). Recently in 2011, bubble-trapping was used to identify origins of replication from synchronized HeLa and lymphoblastoid cell lines using the ENCODE tiling array for a 30-Mb region of the human genome. At this time only the Mesner and Hamlin group has used bubble trapping to identify origins, but their results indicate that the method may be applicable for identifying origins of DNA replication throughout the human genome (Mesner et al., 2011).

**RIP (Replication Initiation Point)**

Replication Initiation Point (RIP) mapping is based on the RNA primers from both the leading and the lagging strand. Mapping the RIP requires careful extraction of total
genomic DNA to maintain the RNA primers. Replication intermediates (RI) are separated using BND cellulose to bind single-stranded regions. Free DNA ends are then phosphorylated using T4 polynucleotide kinase to guarantee exhaustive digestion of all DNA that is not protected by an RNA primer. To prevent contamination with RI-like nicked DNA, lambda exonuclease is used to remove all DNA that is not protected by an RNA primer. DNA primers are then annealed to the RIs and Vent exo-DNA polymerase is used for the primer extension towards the ligated RNA primers. Vent exo-DNA polymerase stalls when it encounters the primed RNA at the 5’end of the DNA strand. The primer extension products are then sequenced (Gerbi & Bielinsky, 1997) to identify the RIP. Vent exo-DNA polymerase is a polymerase with high fidelity that cannot use RNA as a template and stops at RNA/DNA junctions.

RIP provides nucleotide level resolution for the replication initiation point. Abdurashidova, et al (2000) used RIP to map the human lamin B2 origin. Ligation Mediated PCR (LM-PCR) was necessary to enhance the signal which otherwise was weak (Abdurashidova, et al, 2000). The results from the RIP mapping of lamin B2 were impressive. A single band was produced, demonstrating that a precise mapping of the initiation site for DNA replication was possible (Gerbi, S. Personal communication).

A major limitation of RIP is that the region of interest must be defined and have known sequence in order to design the necessary primers for the extension step. In addition, the results must be interpreted with caution when there is an origin cluster or initiation zone, because primers designed within overlapping origins will likely lead to a smear and poorly defined bands. However, once putative plant origins are identified by other techniques, RIP
may be useful for higher resolution mapping and confirmation of putative replication start sites (Figure 4).

**In situ Approaches**

**Fiber Autoradiography.** Initial work by Cairns, et al (1963) used H\(^3\) thymidine incorporation into genomic DNA, which was subsequently extended as a fiber allowing visualization of replicating structures along the entire *E. coli* chromosome. Huberman, et al (1966) adapted the method and visualized the chromosomal fibers in hamster cells. Huberman, et al (1966). (Huberman et al, 1966) also found that some stretches of DNA were longer than the reported lengths of chromosomes, which was due to replicating DNA fibers, due to the relaxed chromatin structure during replication. Later improvements were made using electron microscopy combined with pulse labeling the DNA with H\(^3\) thymidine and autoradiography. By greatly improving the resolution, the first genome wide analysis of replication units in a complex genome (*D. melanogaster*) was possible (Blumenthal, Kriegstein, & Hogness, 1974).

The synthesis of nucleotide analogues bound to biotin (Langer, Waldrop, & Ward, 1981) and the advent of fluorescent microscopy to visualize the DNA fibers led to the development of Fiber- FISH (Fluorescence In Situ Hybridization), which has been extremely valuable for studying newly synthesized DNA. The original Fiber FISH experiments combined the use of ISH (In Situ Hybridization), which used a tritium labeled probe of DNA or RNA, and autoradiography to localize sequences of DNA on chromosomes. Herrick et al (1999) and Jackson et al (1998) refined fiber-FISH by "combing" DNA fibers onto silanized
glass slides, which allowed visualization of the stretched fibers with a mapping resolution of 1 to 4 Kb. Norio, et al (2001) developed a combination of FISH, combing, and mapping of single DNA molecules (Single Molecule Analysis of Replicated DNA, SMARD) using pulse chase labeling with two fluorescent dyes to detect two different halogenated nucleotide analogues. SMARD allowed Norio, et al (2001) to determine the speed and the direction of replication forks within a single molecule of DNA. Importantly, SMARD requires the use of known markers to associate regions of nucleotide analogue incorporation to map locations, and thus is currently not suitable for genome-wide mapping studies. For more on this technique, see Figure 5 and a recent review by Herrick and Bensimon (Herrick & Bensimon, 2009).

**Chromatin Immunoprecipitation (ChIP)**

The use of Chromatin Immunoprecipitation, or ChIP, to map origins depends on antibodies specific for the origin recognition complex (ORC) or other proteins known to be part of the pre-replication complex, such as MCM, RPA and Cdc6 (Costas et al., 2011; MacAlpine et al., 2010; Schaarschmidt, Ladenburger, Keller, & Knippers, 2002; Wyrick et al., 2001). For successful ChIP, the composition of the pre-replication protein complex must be known in order to make the necessary antibodies.

Cells are first fixed with formaldehyde and the chromatin is extracted (Ren et al., 2000). The chromatin is sheared and immunoprecipitated using antibodies against the specific protein that forms part of the pre-replication complex. The crosslink is then reversed to release the bound DNA sequences, which are then amplified, labeled, and hybridized to a genomic microarray or analyzed by deep sequencing (Eaton, Galani, Kang, Bell, &
MacAlpine, 2010). Since there are more ORC binding sites than active origins (DePamphilis et al., 2006), additional ChIP maps with proteins such as members of the MCM complex, which are also associated with active origins, should help to distinguish active from inactive origins and other non-origin sites that bind to ORC. ChIP mapped origins using should be validated by another technique, especially for Arabidopsis or any other species that lacks previously characterized origins.

For origin mapping by ChIP to work well, it is important to have knowledge of the proteins that make up the pre-replicative complex (PRC) of the organism in question, and to be able to produce high quality antibodies for ChIP that are specific for the protein of interest. Other important factors are good fixation of the ORC (or MCM) to the bound DNA along with efficient elution from the beads. To lower the background noise inherent to the ChIP method and increase resolution, Rhee and Pugh (2011) included a step in which lambda exonuclease is added to the immunoprecipitated DNA bound to the resin, protecting it from being digested by the lambda exonuclease (Rhee & Pugh, 2011).

Austin, et al (1999) used ChIP with ORC (DmORC) to identify an origin within the *Drosophila melanogaster* chorion gene locus (Austin, Orr-Weaver, & Bell, 1999). Wyrick et al, (2001) were the first to use ChIP to map the origins of DNA replication for an entire genome (*S. cerevisiae*) by microarray hybridization. The Wyrick, et al (2001) results confirmed previously reported origins and discovered a new subset of origins.

ChIP is now a widely used to study protein-DNA interactions in whole genomes (MacAlpine et al., 2010). It is important to note that well characterized DNA replication machinery was critical for the success of the ChIP-based origin analysis. Even though some
differences have been reported, the core DNA replication machinery is similar among eukaryotes, and similar approaches can be applied to study them as shown in Figure 6 (Costas et al., 2011; Shultz et al., 2007). Thus, we can speculate that information on the replication machinery from other organisms can also be applied for using ChIP to map origins of replication in plants. At this time, no data has been published for mammalian pre-RC ChIP-chip or ChIP-seq because of the poor enrichment over the background (Gilbert, 2010; Schepers & Papior, 2010), but there are new efforts to try to overcome the ChIP background as stated before (Rhee & Pugh, 2011).

**Origin Analysis**

As described above, several different methods can be used to map origins. Determining which method to use depends largely upon the number of putative origins to be analyzed. Depending upon the scale, either PCR or large-scale genomic methods can be used. When small numbers of origins are being analyzed PCR (competitive or qPCR) can be used, but it is necessary to know the location of putative origin in order to design primers. For example, PCR analysis is widely used to examine nascent strands enrichment (Giacca et al., 1994), in addition to being used for confirming putative origins. However, rapid advances in genomic analysis are providing opportunities to efficiently map origins across large regions or entire genomes. Examples of such methods include the use of tiling microarrays (Lucas et al., 2007) and deep sequencing, which is becoming faster and cheaper.

It is important to choose the proper sequencing method due to the differences in length and number of readings that can impact the ability to match the putative origins with
the sequenced genome. Several companies currently provide services for deep sequencing. For example, an Illumina Genome Analyzer II was used to identify putative origins from nascent strands from plants (Costas et al., 2011) and humans (Martin et al., 2011).

Episomes

As described earlier, the first described origin of DNA replication from S. cerevisiae was the ARS (autonomously replication sequence). Inclusion of the ARS is required for extrachromosomal replication of the 2 micron plasmid as well as E. coli plasmids by yeast replication machinery (Stinchcomb et al., 1979). The success with yeast led to similar experiments to test for autonomous replicating sequences in human cells. However, it was found that any sequence of great enough length (10 Kb) supported replication, regardless of the sequence (Caddle & Calos, 1992). Similar studies found that sequences from drosophila, human, or E. coli were competent for replicating extrachromosomally in Drosophila cells (J. G. Smith & Calos, 1995).

Piechaczek et al (1999) developed a plasmid using SV40 in which the T-antigen, that is required for viral replication, was deleted and replaced by a MAR from the human interferon beta-gene. The SV40 replication origin remained functional, and the resulting episome (pEPI-1) was the first example of an episome capable of replication in mammalian cells without requiring a viral protein (Piechaczek, Fetzer, Baiker, Bode, & Lipps, 1999). A subsequent study by Schaarschmidt et al (2004) showed that the replication proteins ORC2 and MCM3 bound to the pEPI-1 episome, and that it continued stable extrachromosomal replication over many cell generations (30 cell generations) (Schaarschmidt, Baltin, Stehle,
Importantly, the MARs that substituted the T antigen in the SV40 plasmid, allowed the plasmid to replicate extrachromosomally but only once per cell cycle. It was shown that pEPI 1 follows the mitotic division of the cells and binds the replication machinery. The finding that a MAR sequence is important for pEPI-1 replication raises the possibility that the system can be used to determine whether other genomic sequences can support episomal replication and increase our understanding of what constitutes an origin of replication in higher eukaryotes.

Researchers continue to develop artificial chromosomes for use in improving plant transgene expression (Ananiev et al., 2009; Birchler, 2010; Dhar, Kaul, & Kour, 2011; Lin, Koo, Zhang, St Peter, & Jiang, 2011). Yu et al, (2007) and Carlson et al (2007) have developed artificial plant chromosomes, the minichromosome B and the MMC respectively. The minichromosome B and the MMC potentially could be used in future studies to understand and address questions about what sequences are important for the minichromosome replication and particularly what sequences act as origins.

**Differences Between Plants and other Higher Eukaryotes**

Okazaki fragment sizes are not known in plants, which may complicate protocols, such as assays based on nascent strand analysis. In these protocols, the size of the DNA fragment to be used is limited so that the smallest fragments isolated are larger than the largest Okazaki fragments, but small enough to be close to the origin of replication.

Much research in DNA replication has made use of synchronized cell cultures. However, the availability of synchronizable plant cell cultures is limited, especially for
*Arabidopsis* and other plant species with sequenced genomes. Two cell lines (MM1 and MM2d) of *Arabidopsis thaliana* Ler can be synchronized using sucrose starvation (Menges & Murray, 2002). However, the genome of this ecotype has not been sequenced, and our unpublished data has shown a substantial amount of copy number variation between the Ler and Col0 ecotypes. There are also more general problems associated with synchronized cell cultures. Synchronization typically requires the use of chemicals or starvation to inhibit replication, increasing the likelihood that artifacts may occur (Anglana, Apiou, Bensimon, & Debatisse, 2003). One alternative is the use of sorted cells, which can enrich for cells in S phase. For experiments requiring nascent strands especially if lambda exonuclease is used, it will be necessary to sort large amounts of material, because for the numbers of S phase nuclei and nascent strands are both low.

While cell cultures have advantages for origin mapping, it would be interesting to compare the origins used in cell culture with the origins used in intact tissues such as root or shoot meristems to determine if origin usage differs in different developmental circumstances. In metazoans, some studies have shown that alternate origins can be used, depending on the developmental program. Some examples include the replication patterns in the embryos of *Xenopus levis*, which changes after transcription starts (Hyrien & Mechali, 1993), and changes in the replication pattern at the *Igh* locus during B cell development (Norio et al., 2005). These experiments were done in cell cultures and it is not known if the same happens in the intact organism. Eventually, the techniques developed for mapping origins in cell cultures will pave the way for mapping origins in intact tissues of whole organisms.
Validation of Putative Origins

Once a putative origin is identified by one method, validation is necessary by a different method. For example, a genome-wide analysis of nascent strand abundance is informative, but can be made much more convincing if a significant number of the putative origins thus identified can be confirmed by 2-D gel analysis (Aparicio, O.M., personal communication). Finally, novel techniques such as SMARD can help discriminate between individual origins and initiation zones, while also providing information about the speed of fork movement (Verma et al., 2011).

Origins of DNA Replication and Chromatin Structure

Transcription and replication require an open chromatin conformation to allow proteins to bind to the DNA. In animal studies, origins have been found in CpG islands, AT-rich elements, and DNA unwinding elements (DUE) as well as DNase sensitive sites and nucleosome free regions, histone acetylation sites, and transcription start sites (TSS) (Cadoret & Prioleau, 2010; Gondor & Ohlsson, 2009; Masai, Matsumoto, You, Yoshizawa-Sugata, & Oda, 2010; Mechali, 2010). Since chromatin remodeling has an effect on both replication and transcription, it is attractive to suggest that a specific sequence may not be required to start replication but, rather, that origin usage depends on chromatin structure and other factors that may be cell-specific.

The eukaryotic genome must be packaged into the nucleus in a manner that is extremely compact, but highly organized to facilitate the critical nuclear processes necessary for the survival of the cell. The DNA is compacted by multiple interactions with histone proteins, which allow the entire genome to fit within the nucleus and control access to
regulatory factors. Chromatin structure and DNA accessibility are controlled during development and differentiation by chromatin remodeling, a dynamic process that involves specific histone and DNA modifications and chromatin remodeling protein complexes (O. Bell, Tiwari, Thoma, & Schubeler, 2011).

To understand how chromatin structure is controlled, it is necessary to know the basic levels of chromatin structure (Figure 7). The diameter of the double helix DNA ranges from 2 to 2.3 nm. The DNA is wrapped twice around a histone octamer (nucleosome), which has a diameter of 11 nm. The nucleosome is composed of 2 molecules of each core histone (H2A, H2B, H3 and H4). The presence of the histone H1 links the nucleosomes, further packaging the DNA, resulting in the formation of the 30-nm fiber. Higher-order packaging beyond the 30-fiber occurs, but details are still largely unknown (Fussner, Ching, & Bazett-Jones, 2011; Maeshima, Hihara, & Eltsov, 2010). Current models postulate that the 30-nm chromatin fibers form ~100-nm loops in which the base of the loop is attached to a matrix or scaffold (Flors & Earnshaw, 2011). During prophase, each series of 100-nm loops are coiled to form the chromatid.

Genomic DNA is most tightly packaged during mitosis, when chromatids separate to ensure that the daughter cells each have a complete set of chromosomes (Bloom & Joglekar, 2010). Chromatin structure is re-organized during interphase. During replication and transcription, the DNA becomes accessible to replication machinery, transcription factors, RNA polymerases, and other auxiliary factors. However, much of the genome remains condensed throughout most of interphase. This fraction is usually referred to as heterochromatin, which can be either constitutive or facultative.
Each core histone within the nucleosome can be modified. Initial studies of histone modification (Tamaru et al., 2003), found that various lysines within the histone H3 tail could be methylated or acetylated, the serines phosphorylated, and the arginines methylated. Various histone modifications can have very different effects on chromatin structure, with modifications that appear subtle sometimes having profound effects on accessibility. For example, in *Arabidopsis thaliana* H3K9me1 and H3K9me2 are considered epigenetic marks for inaccessible (inactive) chromatin and are frequently associated with aggregated clusters of heterochromatin called chromocenters (Fransz & de Jong, 2011), whereas the trimethylated version of the same residue - H3K9me3 - is associated with accessible (active) chromatin (Berger, Dubreucq, Roudier, Dubos, & Lepiniec, 2011) [and reviewed in (C. Liu, Lu, Cui, & Cao, 2010)]. The large number of histone post translational modifications (PTMs) and the fact that they can occur in a bewildering variety of combinations suggest that chromatin may assume a wider variety of different conformations associated with its various functions (Garcia et al., 2007).

The majority of studies on histone PTMs are concerned with the relationship between a given modification (or epigenetic “mark”) and transcriptional activity or gene expression. However, histone PTMs during chromatin remodeling is also an important process that controls DNA replication.

A very important question that remains to be answered is how are these histone modifications (epigenetic marks) transmitted through multiple cell generations? It is of vital importance to pass the correct epigenetic information to both daughter cells following DNA replication and cell division. Of particular importance in this process is that the epigenetic
marks on the chromatin must be reconstituted after the passing of a replication fork. How is this achieved? In the paragraphs that follow, I will give a short overview of processes currently thought to be involved in passing epigenetic marks to newly replicated DNA.

The proliferating cell nuclear antigen (PCNA) has been reported to have an important role in maintenance of the methylated DNA through the binding of DNA methyltransferase 1 (DNMT1) (Chuang et al., 1997; Hermann, Goyal, & Jeltsch, 2004). In addition to DNMT1, PCNA maintains histone PTMs on newly replicated DNA by binding the chaperon chromatin assembly factor 1 (CAF1) (Shibahara & Stillman, 1999; Zhang, Shibahara, & Stillman, 2000). CAF1 coordinates nucleosome assembly by adding new H3 and H4 (S. Smith & Stillman, 1989). Another chaperon involved in the transmission of histone PTMs is the anti-silencing function 1 (ASF1), which plays a central role in re-establishing histone PTM patterns on newly replicated DNA. During replication fork passage MCM2-7 is thought to disrupt nucleosomes and evict the histones. The chaperone ASFS1 binds H3-H4 and minichromosomal maintenance 2-7 (MCM2-7) (Groth et al., 2007). The MCM-(H3-H4)-Asf1 complex recruits histone acetyl transferase and transfers the newly acetylated H3-H4 to the CAF1 and Rtt106 chaperones for incorporation into newly synthesized DNA (Avvakumov, Nourani, & Cote, 2011). Newly synthesized DNA transiently carries the acetylated marks H4K5ac and H4K12ac deposited by CAF1, which are related to new histone deposition (Jasencakova, Meister, Walter, Turner, & Schubert, 2000; Mayr, Jasencakova, Meister, Schubert, & Zink, 2003; Aline V. Probst, Dunleavy, & Almouzni, 2009). These acetylated marks are then removed and the proper modification takes place.
*Arabidopsis thaliana* has two *ASF1* genes (*AtASF1a* and *AtASF1b*). Mutation of either one alone causes no effect whereas a double mutation (*Atasf1a*/*b*) severely inhibits growth and development (Zhu et al., 2011). The importance of ASF for DNA replication is clearly observed in *A. thaliana*, as the *Atasf1a*/*b* double mutant has stalled replication forks and activated DNA repair (Zhu et al., 2011).

Following replication, it is critical that newly synthesized DNA be reconstituted into chromatin to maintain genome integrity. With the recent development of the nascent chromatin capture technique, it has become possible to study the dynamics of the histone PTMs that occur on newly replicated DNA (Sirbu et al., 2011). Therefore, we will be able to understand what happens immediately after the replication fork has passed during DNA replication.

Next, I will describe how it is suggested that the heterochromatic PTMs are maintained epigenetically. How euchromatin PTMs are inherited is another story that is not clearly understood yet.

**Replication of Centromeres and Pericentromeric Heterochromatin**

The centromere and pericentromere regions are very important because they contribute to the correct chromosome segregation during mitosis. They are form of constitutive heterochromatin (Aline V. Probst et al., 2009). The centromeres regions are not specified by primary sequence, except in *S. cerevisiae*, (Karpen & Allshire, 1997), but instead are determined by epigenetic marks (Allshire & Karpen, 2008). The common epigenetic mark for centromeres is the histone H3 variant CenH3 (Boyarchuk, Montes de Oca, & Almouzni, 2011; Aline V. Probst et al., 2009), while the pericentromeric area is
recognized as constitutive heterochromatin with the epigenetic mark HP1 (A. V. Probst & Almouzni, 2011). HP1 binds H3K9me3 at the pericentromere (Bannister et al., 2001; Boyarchuk et al., 2011; Lachner, O'Carroll, Rea, Mechtler, & Jenuwein, 2001; A. V. Probst & Almouzni, 2011). The histone methyltransferase Suv39h (suppressor of variegation), which methylates H3K9, binds to HP1 (Aagaard et al., 1999) self-propagating to the neighboring histones and perpetuating itself (Grewal & Moazed, 2003).

DNA and histone modifying complexes interact with the replication fork machinery to quickly re-establish heterochromatic structure following centromeric replication. The histone modifying complex includes histone methyltransferase (Suv39h), CAF-1, and putative histone deacetylases (Maison, Quivy, Probst, & Almouzni, 2010). Quilvy et al (2004) originally demonstrated that during mid/late S phase and the pericentromere is replicated, CAF1 and HP1 bind to PCNA (Quivy et al., 2004). This suggested a model where CAF1 deposits de novo H4K5ac and H4K12ac and also delivers HP1 (Maison et al., 2010). Loyola et al (2009) demonstrated that CAF1-HP1 forms a complex with the methyltransferase SetDB1 that methylates H3K9me, a substrate for the methyltransferase Suv39h to produce H3K9me3 (Loyola et al., 2009).

To ensure proper spatial and temporal gene regulation following DNA replication, the chromatin needs to recover its original structure accomplished by the retention of repressive chromatin structures, mediated by the Polycomb group of proteins (PcG). Recent studies have shown that polycomb (PcG) proteins remain bound to the replicating DNA (Margueron & Reinberg, 2010). Methylation of H3K27me3 is mediated by polycomb repressive complex 2 (PRC2), which also binds its target H3K27me3. It has been
demonstrated that PRC2 that catalyses H3K27me3 methylation binds to PCNA during S phase (Hansen et al., 2008), suggesting that after the replication fork passes, PRC2 binds to the parental H3K27me3 and maintains the mark in proliferating cells (Blomen & Boonstra, 2011).

Non-coding RNA (ncRNA) has also been demonstrated to play a role in heterochromatin formation. It was shown recently that RNA interference (RNAi) influences heterochromatin propagation and heterochromatin inheritance. The siRNA bound to argonaute, cleaves the mRNA in the cytoplasm. Then, RdRP (RNA dependent RNA polymerase) binds to the cleaved mRNA and produces secondary siRNA. The secondary siRNA binds to another argonaute (NRDE-3) and transports it to the nucleus, where it interacts with the RNA pol II that is transcribing the original mRNA. The siRNA-NRDE-3 binds to the newly transcribed mRNA, depositing H3K9me3 at the same time in the gene being transcribed. The H3K9me3 pattern is inherited by two generations and then decreases (Zaratiegui & Martienssen, 2012). A similar mechanism has also been found in plants (Volpe & Martienssen, 2011).

**Research Rationale**

The focus of my research has been to characterize potential origins of DNA replication on the distal long arm of *Arabidopsis thaliana* Chr 4, and to determine the prevalence of some prominent posttranslational histone modifications as cell cycle proceeds through G1, S and G2/M. The DNA replication origin is a site from which the cell starts to duplicate its DNA to make an identical copy. In a eukaryotic cell, DNA replication starts more or less simultaneously at multiple points within the genome, which is necessary to
ensure complete replication in a timely manner. As noted previously, the sequences that serve as origins of DNA replication in higher eukaryotes, and the regulation of origin usage, are poorly understood, especially in plants.

Newly synthesized DNA (nascent DNA strands) have been used widely to map origins of DNA replication. The complete protocols for isolating and mapping nascent strands are described in Chapter II. I attempted two protocols for the isolation of nascent strands, which included size selection (Neat SNS) and lambda exonuclease (Exo SNS). The isolated nascent strands were hybridized to a custom tiling array for Chr 4 of Arabidopsis thaliana in an effort to map for origins of DNA replication.

To date, no specific consensus sequence has been identified as an origin of DNA replication in higher eukaryotes. The lack of a specific DNA sequence supports the idea that origins of DNA replication are regulated by chromatin structure, with open chromatin allowing access to replication proteins. It is now well known that PTM of histones influence the chromatin structure.

To understand the role of key well-studied histone PTMs, we examined the change in the levels and subnuclear localization that occur during cell cycle progression. The nuclei were first sorted by fluorescence activated cell sorting (FACS) as outlined in Chapter III into G1, S and G2/early M phases. The modified histone levels in the G1, S and G2/early M phases nuclei were then measured by quantitative Western blot analysis, followed by an immunolocalization to compare the localization patterns for specific modified histones.
REFERENCES


Anachkova, B., & Hamlin, J. L. (1989). Replication in the amplified dihydrofolate reductase domain in CHO cells may initiate at two distinct sites, one of which is a repetitive sequence element. Mol Cell Biol, 9(2), 532-540.


Figure 1. Origin of DNA replication and nascent strands. A) The pre-RC at the origins of DNA replication. b) Replication bubble with nascent strands. c) Newly synthesized DNA can be enriched by immunoprecipitation of BrdU pulse-labeled DNA, treatment with lambda exonuclease, or binding to BND cellulose.
Figure 2. Two dimensional (2-D) gels. Modified from (Brewer and Fangman, 1987) 2-D gels allow the separation of replicative forms of DNA. The first dimension separates by mass while the second separates by structure. A) Fork progression. B) The bubble formed at the initiation of replication can be observed by its migratory pathway. C) forks converging. D) asymmetric bubble and fork progression.
Figure 3. Matrix Attachment Regions (MARs). MARs plays an important role in the nucleus. To isolate MARs nuclei preparation are required, followed by removal of histones to make halos, then elimination of unbound DNA fragments and removal of matrix proteins.
Figure 4. Replication Initiation Point (RIP) mapping. This method uses nascent strands and primers to map origins at a nucleotide level. The Transition Point (TP) between nascent strands and Okazaki fragments is mapped with primers on both strands located close to the origin. The origin sequence will be located at the TP and it is solved by sequencing gels.
Figure 5. Single Molecule Analysis of Replicated DNA (SMARD). Modified from figure 1b (Norio et al., 2005). The SMARD assay provides information like fork direction and fork speed. Cells are sequentially labeled with two different thymidine analogues (IdU, Iodine deoxy-uridine and CldU, chloro deoxy-uridine). After the incorporation, the DNA is extracted and fragmented with restriction enzymes. The fragments are stretched into microscope slides and hybridized in situ with a probe to identify the desired fragments, prior to mapping IdU and CldU incorporation with antibodies conjugated to different fluorophores.
Figure 6. Chromatin Immunoprecipitation (ChIP). The replication machinery binds origins to start replication. Putative origins can be isolated by immunoprecipitation with high quality antibodies against replication proteins. ORC and MCM proteins interact with the origin of DNA replication making ChIP for ORC and MCM complexes ideal to map origins.
CHAPTER II

Mapping the Origins of DNA Replication from the Distal Long Arm of Arabidopsis thaliana Chromosome 4
All the statistical analyses in chapter II was performed by Dr. Peter E. Pascuzzi
DNA replication is a fundamental process that all living organisms must be able to perform to pass their genetic information on to subsequent generations. Despite the importance of DNA replication, much of what we currently know comes from studies of model organisms such as viruses, bacteria and *Saccharomyces cerevisiae*. The recent advent of novel molecular technologies have enabled studies of entire genomes and accelerated our efforts to understand DNA replication in more complex higher eukaryotes such as multicellular fungi, plants and animals.

The high degree of conservation of the replication components allows researchers to apply the protocols and models developed from studies of simpler eukaryotes to develop models of DNA replication in higher eukaryotes, such as plants. While the general process of replication is highly conserved, the role of the components of the replication machinery may vary from organism to organism.

Much of our current understanding of DNA replication initiation is based on the “replicon model” as originally proposed by Jacob and Brenner (Jacob & Brenner, 1963). The replicon model postulates that a trans-acting factor (initiator) recognizes a specific genetic element (replicator) to initiate DNA replication of a defined unit (replicon) of the genome. The model proposes the existence of a regulator (protein) for DNA replication that recognizes a specific site (origin) to start replication.

As I noted in Chapter 1, the sequence of DNA replication origins in higher eukaryotes is poorly conserved, which makes simple modeling difficult. Thus far, the best characterized
origins of DNA replication for an entire genome (Gilbert, 2010) are from either well-studied bacteria (*Escherichia coli*), budding yeast (*Saccharomyces cerevisiae*) (Wyrick et al., 2001), or fission yeast (*Schizosaccharomyces pombe*) (Dai et al., 2005). Recently Costas, et al. (2011) developed a map of putative genomic origins of DNA replication for *Arabidopsis thaliana* (Costas et al., 2011), but none have been tested thus far to validate whether the identified sequences serve as true DNA replication origins.

As discussed in chapter I, the first origin in eukaryotes (ARS1) was found using the 2 micron plasmid of *S. cerevisiae* (Brewer & Fangman, 1987). Once the ARS were discovered its use has allowed the development of a database “OriDB” that includes all ARS origins reported to date (http://www.oridb.org).

Most studies of DNA replication origins in higher eukaryotes have relied on cell culture lines, which are considered to be more homogeneous cell populations that can yield high amounts of sample material, necessary for some analyses. For example, CHOC 400, a variation of the Chinese Hamster Ovary cell line (CHO) created in Hamlin’s lab (http://people.virginia.edu/~jlh2d/), has approximately 1000 copies of a 132 kb fragment containing the DHFR gene (Dihydrofolate reductase), which facilitated the discovery of the first replication origin in metazoans within the DHFR amplified region (Heintz & Hamlin, 1982).

The DHFR origin of replication has served as a positive control to confirm additional origins. These include lamin-B2 (Abdurashidova, Riva, Biamonti, Giacca, & Falaschi, 1998), beta-globin (Kamath & Leffak, 2001), c-myc (Waltz et al., 1996) and rDNA (Little, Platt, & Schildkraut, 1993).
Several of the origin mapping procedures are based on the use of short nascent strands (SNS). Small nascent strands are produced from the leading strand and have an RNA primer cap at the 5’ start site of the SNS. The lambda exonuclease procedure takes advantage of the RNA primer to protect the SNS while unprotected DNA strands are digested by the nuclease. A recent study (Karnani, Taylor, Malhotra, & Dutta, 2010) mapped 150 origins of replication from HeLa cells using lambda exo SNS, while a second study (Cayrou, Coulombe et al. 2011) mapped 2748 origins from mouse chromosome 11 and 6184 origins from the Drosophila genome. These promising origin-mapping results for complex genomes indicate that a similar approach could be used for plants, especially ones with smaller genomes such as Arabidopsis.

Unfortunately mapping origins of replication in plant genomes has proven to be more difficult, which may be due to several factors unique to plants such the presence of a cell wall or higher levels of nucleases. In addition, the actual size of plant Okazaki fragments is unknown, which makes the SNS size-selection less precise. In contrast to animals, there are no well-defined plant origins of replication that can be used as positive controls. The lack of a well-defined plant origin of replication makes the analysis and validation more difficult, making it necessary to use more than one method to compare common regions. An additional method that has been used for mapping genomic origins of replication is chromatin immunoprecipitation (ChIP) using antibodies to the proteins known to be components of the replication machinery during initiation. Examples include the use of ORC2 ChIP to map fly (Sciara coprophila) origins (Bielinsky et al., 2001) and ORC and MCM ChIP to map origins in S. cerevisiae genome (Wyrick et al., 2001). However, the use
of ORC or MCM ChIP to map origins in higher eukaryotes has been more complex (Gilbert, 2010). Gilbert (2010) noted that to date there are no published examples of ChIP mapped origins in higher eukaryotes, perhaps due to low levels of enrichment above background.

However, recently Eaton, et al. (2011) used ChIP to map ORC binding sites in three Drosophila cell lines and developed a support vector machine (SVM) to predict ORC binding sites and origins of replication (Eaton et al., 2011). Unfortunately, whether the predicted origins are actually active, remains to be determined. Costas, et al (2011) used 5-bromo-2′-deoxyuridine (BrdU)-labeling of nascent DNA in combination with ORC1 and CDC6 ChIP to characterize origins of replication in an A. thaliana suspension cell line. The ORC1 and CDC6 binding was extensive and required comparison with the BrdU mapping data to reduce the number of binding sites to approximately 1,500 putative origins.

In our study, we also used BrdU labeling of nascent strands, but found that the method is susceptible to random nicking of DNA (Hamlin et al., 2010; Kumar et al., 1996) that results in an unacceptable level of background signal (noise) when hybridized to a microarray. Our results showed that the high level of background for each of three biological replicates (bioreps) made the correlation too low to be interpreted. Because of the lack of reproducibility with the BrdU SNS method, we used two alternative SNS methods. The SNS Neat (we gave the name Neat to say without BrdU, like a drink without ice) protocol uses isolated nascent strands fractionated by size. The goal of the SNS Neat method is to select the short nascent strands by using a size range, which reduces the potential contamination by Okazaki fragments while increasing the resolution by removing larger fragments that are far from the origin (Kumar et al., 1996). Our third SNS protocol treats the
isolated SNS with lambda exonuclease (Giacca, Pelizon, & Falaschi, 1997; Sibani, Price, & Zannis-Hadjopoulos, 2005), which digests any DNA that is not protected by an RNA primer, followed by size selection to remove the Okazaki fragments. In each protocol, the SNS fractions are labeled and hybridized to a custom-designed NimbleGen microarray for chromosome 4 (AtChr4) of Arabidopsis thaliana Col-0. AtChr4 is the smallest of the five chromosomes in Arabidopsis, which allows greater microarray resolution and duplicate data from each microarray. AtChr4 has been well-studied, and its replication timing and histone modification patterns are known from our previous work with this cell line (Lee et al., 2010).

We found that the SNS Neat correlation between the three biological and the three technical replicates was the highest of the three SNS methods used in our study. A comparison made between Neat and Exo SNS data showed 207 common peaks for the two methods. The Neat and Exo SNS were queried using the annotations in TAIR10. The only similarity that was uncovered is their depletion in transposable elements (TEs). Their AT content is not different than the average for the distal long arm (DL) of chromosome 4. While Neat SNS peaks are poorer in AT content and show a preferred location in gene bodies, Exo SNS peaks are slightly AT rich and enriched for unannotated and the 3’ untranslated region (UTR3). When compared with replication timing results, both data sets did not show any enrichment for early replicating zones, but Neat SNS showed enrichment for mid and late replicating regions.
MATERIALS, METHODS AND NOTES

Cell Culture

The *Arabidopsis thaliana* cell line (Col-0, ecotype Columbia), which was kindly provided by Dr. Iris Meier (Ohio State University), was grown and maintained as previously described by Lee, Pascuzzi, et al (Lee, Pascuzzi et al. 2010). To ensure that the cells were actively replicating DNA, a 7-day split protocol was used as previously described by Lee, Pascuzzi et al (2010) but without the paraformaldehyde fixation treatment. The ‘7-d split culture’ was prepared and the split culture was grown for an additional 17 h and harvested. For each experiment the two flasks from the ‘7-d split culture’ represent one biological replicate.

Genomic DNA Isolation

Genomic DNA was extracted using the CTAB protocol described in (Murray and Thompson 1980;(Allen, Flores-Vergara, Krasyanski, Kumar, & Thompson, 2006) Two flasks (50-mL each) were used for the isolation of genomic DNA from the frozen cells of the 7-d split culture cells. Seven to 8 grams of frozen cell pellets were isolated from each flask. Approximately 1 g of the dry, frozen cells was weighed in a 2-mL amber micro centrifuge tube. One bearing ball and 500 µL of cold 70% ethanol were added to each tube to and the mixture was ground twice in a Retsch mixer mill MM301 at a frequency of 25 Hz for 2 minutes. The ethanol was removed by centrifugation at 13,400 x g for 1 minute. The CTAB (1,200 µL) with 0.5% v/v β-mercapto-ethanol was added to the cell pellet, mixed gently, and incubated at 65°C for 45 minutes with gentle rotation. The samples were then centrifuged at
13,400 x g for 5 minutes to remove the cell debris. The supernatant was collected in a fresh 2-mL micro centrifuge tube and 800 µL of phenol-chloroform-isoamyl-alcohol was added. The samples were mixed gently for 10 minutes in order to prevent DNA strand breakage. The samples were then centrifuged and the upper layer was collected and transferred to a fresh 2 mL tube. Chloroform (800 µL) was added and mixed gently for 10 minutes and then centrifuged at 13,400 x g for 5 minutes. The supernatant was transferred to a fresh tube with 1000 µL of cold isopropanol. The DNA pellet was washed with 70% cold ethanol and air dried at room temperature in the dark for 1 hour. The centrifuged DNA was then resuspended into 25 µL of TE (Tris EDTA pH 8) buffer. The entire procedure was completed with minimum exposure to light to avoid nicking of the DNA containing the BrdU.

**Size Fractionation by Alkaline Sucrose Gradient**

Before preparing the gradient, the DNA was measured using NanoDrop 2000 from Thermo Scientific. The DNA concentration was adjusted to 20 µg of DNA in 500 µL per gradient using 2 N NaOH and 0.5 M EDTA. The gradient was made by adding 9 layers of 500 µL each layer, ranging from 10% to 30% sucrose, to a 5-mL centrifuge tube. The final layer was topped with 500 µL of the prepared DNA. The gradient tubes were centrifuged for 15 hours at a speed of 260,000 x g at 10°C in a Beckman Optima ultracentrifuge equipped with a swinging bucket rotor SW 55. Two hundred µL fractions of the gradient was then sequentially removed and collected in a 1.5-mL amber microcentrifuge tube. Two hundred µL of Tris HCl buffer was then added to each tube and mixed. A total of 26 fractions corresponding to one gradient were collected.
BrdU Short Nascent Strands (BrdU SNS)

We used the procedure described by Giacca, et al (1997) with minor modifications. BrdU incorporation was done using a 7-d split culture, the 7-d split culture was grown for 16 h and then labeled for 1 h with 100 mM BrdU (Sigma B9285), as described by Lee, Pascuzzi, et al (2010).

An aliquot was taken from each fraction of the alkaline sucrose gradient to analyze BrdU incorporation. Each aliquot were boiled for 5 minutes and immediately placed in ice. A Duralose-UV membrane (Stratagene) was then prepared and a grid of 1 cm² was drawn for each fraction sample. One μL from each fraction was then spotted, allowed to dry, and an additional 1 μL was spotted onto the same spot. The membrane was then dried in the dark for 5 to 10 minutes and cross-linked twice at optimal setting (Stratagene). The Duralose-UV membrane was placed in Blocking buffer for 1h for fluorescent Western blotting (Rockland Inc. cat # MB-070). Anti-BrdU (Invitrogen) antibody was then added (1:10,000 dilution) and the membrane was gently shaken overnight at 4°C. The membrane was then rinsed with Tris buffer saline (50 mM TrisHCl pH7.4, 150 mM NaCl) plus 0.1% tween 20 (1xTBST) three times for 10 minutes at room temperature. Anti-mouse secondary antibody diluted 1:10,000 in Blocking buffer was added, and the membrane was incubated for one hour at room temperature.

The membrane was then rinsed three times with 1 x TBST for 10 minutes at room temperature and scanned with an Odyssey (LI-COR, Lincoln, NE USA) according to manufacturer’s instructions. The signal was quantified using the Odyssey software (version 2.1). An aliquot of each fraction was then run in a 1% alkaline agarose gel to determine the
size range for each fraction. The fractions containing BrdU-labeled DNA in a 1 to 3 Kb size range were selected and pooled for immunoprecipitation.

**Immunoprecipitation of BrdU Labeled SNS**

The immunoprecipitation was done as described in (Lee et al., 2010) except that the DNA eluted with glycine was directly precipitated with 1M Tris pH 8.0, 3 M sodium acetate and 200 proof ethanol without the use of either proteinase K or phenol:chloroform. The eluted and cleaned DNA from the immunoprecipitation was used as template for random amplification, prepared for microarrays according to the manufacturer’s protocols and used for hybridization to the NimbleGen microarray. The flow diagram for the BrdU SNS protocol is shown in Figure 1.

**Neat Short Nascent Strands (Neat SNS)**

We used the protocol described by (Kumar et al., 1996) with minor modifications. DNA was extracted from a 7-day split culture as previously described. The DNA was then fractionated in an alkaline sucrose gradient and the size range was determined by alkaline gel electrophoresis. Fractions that contained the DNA fragments within 1 to 3 Kb were pooled together for subsequent amplification and hybridization to the NimbleGen microarray according to the manufacturers protocols (Figure 2).

An additional control was also included in which the fractions of the gradient containing the larger fragments of DNA were used as a reference for the microarray hybridization. The additional control helped to reduce the signal caused by the small fragments in contrast to when the total DNA preparation was used as a reference. The
reference sample was sonicated to obtain small fragments of DNA to amplify and hybridize to the microarrays.

**Lambda Exonuclease Short Nascent Strands (Exo SNS)**

We used a modified protocol as described by Sibani, et al (Sibani et al., 2005). DNA was extracted from a 7-day split culture using the CTAB protocol as described previously with precautions taken to avoid RNase contamination. The isolated DNA was passed 10 times through a syringe with a 26G3/8 needle and cleaned using a Qiagen column from the Plant DNeasy Kit. The eluted DNA was dissolved in water to a final concentration of 20 µg with a maximum volume of 40 µL, and frozen at -20°C for 24h. The sample was thawed on ice, and an aliquot was spiked it with a linear plasmid control to determine the effectiveness of the lambda exonuclease digestion. The samples were boiled for 5 minutes and rapidly cooled on ice. Ten units of T4PNK (NEB) in 1x T4 ligase buffer and 1 mM ATP was added to the reaction, and the samples were phosphorylated for 30 minutes at 37°C. The kinase reaction was stopped by heat inactivation at 100°C for 3 minutes. Then 15U of lambda exonuclease (NEB) was added and supplemented with buffer to obtain a final volume of 80 µl at a pH ≥ 8.5 for maximal exonuclease activity. Digestion with lambda exonuclease continued for 12 hours at 37°C. The digestion was inactivated at 75°C for 10 minutes. RNase A (Invitrogen) was added to a final concentration of 0.01 µg/µL to remove RNA and the samples were heat denatured. The samples were run in a 1.5% agarose gel with orange G loading dye. The gel was stained with methylene blue (0.02% w/v) and the smear ranging from 1 to 3 Kb was extracted and cleaned with QIAquick Gel extraction Kit (Qiagen). The
eluted DNA was frozen at -80˚C for subsequent amplification and hybridization to the NimbleGen microarray according to the manufacturer’s protocols (Figure 3).

**Labeling and Microarray Hybridization**

For the SNS BrdU hybridization, sheared total genomic *Arabidopsis thaliana* DNA was used as the reference. The reference and the sample were amplified, according to Lippman, et al (2005). After amplification the sample and the reference were labeled with either Cy3 or Cy5 fluorescent dye and hybridized to the first generation Washington State University (WSU) custom-designed NimbleGen array. The NimbleGen microarray was later redesigned by North Carolina State University (NCSU) team for its use in subsequent Neat and Exo SNS experiments.

Karnani, el al (Karnani, Taylor, Malhotra, & Dutta, 2007) have shown that some DNA fragments are replicated asynchronously in human cells. Therefore we decided to use the three gradient fractions containing the largest DNA fragments as the input reference, which were pooled, sonicated and treated as a single input reference. The SNS (target DNA) and input DNA (reference DNA) samples were amplified twice (O'Geen, Nicolet, Blahnik, Green, & Farnham, 2006) using the WGA2 kit (SIGMA), purified, and concentrated using a QIAquick PCR Purification Kit (QIAGEN). Each amplified DNA sample and reference was labeled with either Cy3 or Cy5 fluorescent dye and hybridized to the NCSU custom-designed NimbleGen microarray for AtChr4. Each SNS experiment included six microarrays, which represented three bioreps with the corresponding dye swaps. Sample labeling, microarray hybridization and washing were performed according to the NimbleChip Arrays user’s guide.
Hybridized microarrays were scanned using a PerkinElmer ScanArray Express scanner and quantified using GenePix Pro software v 6.01.

**NCSU Microarray Design and Analysis**

The NimbleGen microarray platform was custom designed for the chromosome IV of *Arabidopsis thaliana* Col-0. The total oligo probe number on the array is 392,714, from which 174,978 probes spotted in duplicates are *A. thaliana* chromosome IV specific. The rest are controls and mismatches. The probes are designed to be isothermal with a length range 50 to 75 bp.

The microarrays were scanned using a GenePix 4000B scanner. The analysis included three bioreps for the BrdU SNS whereas the Neat SNS and Exo SNS included three bioreps in addition to three dye swaps, which served as three technical replicates. GenePix files where formatted as text files using NimbleScan software v. 8 (Roche). The data from the text files was then LOESS and quantile normalized in Limma. Limma is an R package available in the Bioconductor project (http://www.bioconductor.org/). The LOESS and quantile ratios obtained by Limma were then exported back to NimbleScan to determine the SNS peaks. An analysis of several windows was used to determine optimal window size and cut off. The peaks were determined using a sliding window size of 1000 bp with 7 probes above the sliding threshold. Dr. Pete Pascuzzi (NCSU, Molecular and Structural Biochemistry Dept.) developed a custom peak analyzing script.
Notes

To map origins of DNA replication using small nascent strands, it is important that the genomic DNA be free of nicks and double-strand breaks. In our initial experiments BrdU was used to label the nascent replicated DNA, which was size-selected and followed by anti-BrdU precipitation. The BrdU-labeled DNA was then labeled and probed with our AtChr4 NimbleGen tiling array. After repeating the experiment three times, we found that the BrdU-enriched regions were not consistent between the replicated experiments. While the reason for the inconsistency remains unknown, it is known that BrdU increases strand breakage and is frequently used to sensitize DNA for UV damage and repair studies (Gerbi, 2005; Polo, Roche, & Almouzni, 2006; Zeitlin et al., 2009). Random breakage due to BrdU sensitivity may be one explanation, although care was taken to avoid the light during all of the DNA isolation steps.

It is important to note that previous origin mapping efforts have focused on limited regions, such as the DHFR locus, which has DNA that has been amplified by many rounds of selection. While the origins of DNA replication from the amplified DHFR locus served as an excellent model, which is easily mapped, it does not require the high sensitivity necessary for mapping unique chromosomal origins due to the high copy number of the DHFR region. With the advent of genomic mapping it has been necessary to achieve much higher sensitivity or to examine smaller genomes. The time required for BrdU SNS, in addition to the potential problems due to the light-sensitivity of BrdU-labeled DNA, have led researchers to adopt alternative methods for using small nascent strands to map origins of DNA replication such as the use of BND cellulose and λ-exonuclease (Gerbi, 2005).
The λ-exonuclease SNS procedure omits the use of BrdU and relies on an RNA primer to protect nascent DNA from the 5’ to 3’ exonuclease that is used to remove nicked DNA. Therefore, care must be taken to avoid RNases to keep RNA primer cap intact and protect the nascent strands from λ-exonuclease digestion. Another important consideration for the λ-exonuclease procedure is complete digestion of the unprotected DNA. If the unprotected DNA is not completely removed, it will cause a false signal when the final mapping analysis is done. Recently Cayrou, et al (2011) first size selected nascent strand population that was then digested with a specially prepared λ-exonuclease to thoroughly remove the contaminating, unprotected DNA to map the origins of DNA replication in Drosophila and mouse cells (Cayrou et al., 2011). The Cayrou, et al (2011) study has caused controversy because the results raise the possibility that earlier studies that relied on a different source of λ-exonuclease may have been subject to a high level of contamination from undigested DNA, which would have impacted the results of these studies. In addition, Cayrou, et al (2011) showed that the amount of small nascent strand DNA that remains following digestion with the specially prepared λ-exonuclease is vanishingly small (<20 ng per $10^8$ nuclei), demonstrating the challenges faced by researchers who use this method.

In our experiments, we sheared total genomic DNA to have complete digestion during the incubation time with λ exonuclease. The small volume of the DNA sample made sonication difficult. An attempt to shear the DNA with a nebulizer was also unsuccessful because additional handling was required, which resulted in sample loss. We found that the best way to shear the DNA to a size range of 1 to 3 Kb was to pass it through a needle as described above. I think we would have achieved better digestion if size selection was done
before digesting with λ exonuclease as described by Cayrou et al (2011). ChIP-chip analysis v2.0, which was modified to include DyeSaver2 coating reagent (Genisphere) to minimize oxidation of Cy5 (Lee et al., 2010).
RESULTS

BrdU Labeled Short Nascent Strands (BrdU SNS)

BrdU is a thymidine analogue, so when the newly synthesized DNA is being produced, the polymerases will incorporate BrdU in the process. The BrdU can also be incorporated during DNA repair. To obtain short nascent strands, we gave a 60-minute pulse of BrdU to our cell culture, isolated DNA very carefully to minimize nicking, and separated the DNA molecules by size with an alkaline sucrose gradient. DNA molecules ranging from 1 to 3 Kb were immunoprecipitated for BrdU. The 1-3 Kb BrdU-labeled DNA molecules were labeled with Cy5 and hybridized to a tiling array for chromosome 4 of Arabidopsis thaliana. Figure 4 shows that the correlations between the bioreps from the BrdU SNS origin mapping experiment are poor, ranging from a low of 0.036 between bioreps 1 and 3, to a high of 0.288 between bioreps 1 and 2. The low correlation and the high level of background make the results difficult to analyze. It is likely that much of the background noise is due to breakage in the regions of BrdU incorporation at sites of DNA breakage undergoing repair (Gerbi, 2005). While precautions were taken to keep the BrdU-labeled DNA in dark conditions, even limited light exposure could potentially cause breakage of BrdU labeled regions. Regardless, the poor correlation between the three bioreps of our BrdU SNS experiment made further analysis futile.
**Neat Short Nascent Strands without BrdU (Neat SNS)**

Since the procedure of BrdU labeling newly replicated DNA resulted in a very low correlation, which may have been due to high background noise, we decided to isolate nascent strands based on size alone. We selected DNA molecules that ranged from 1 to 3 Kb in an alkaline sucrose gradient. We called this procedure “neat” SNS because it does not include the use of BrdU. In sharp contrast to the BrdU SNS results, the neat SNS procedure yielded the highest correlations between bioreps and technical replicates seen in all the different SNS procedures used in my study (Figure 5). The correlations between bioreps range from 0.91 for bioreps 1 and 2 to 0.93 for bioreps 1 and 3 as shown in Figure 5b. However, Figure 5 also shows that the correlation between technical replicates was low in the same experiments. The technical replicates in this case involved a dye-swap, and it is likely that the low correlations reflect low signal intensity in the red channel (Figure 7).

To analyze all the bioreps and the technical replicates, the contrast method was used to adjust for the differences between the replicates. After the differences were adjusted, we treated them as one dataset and obtained ratios for sample to reference. The ratios (sample/reference) were used in NimbleScan peak finding software to find peaks, using a 1000 bp sliding window and requiring 7 probes to be above the sliding ratio threshold. Using this method, we identified 720 SNS Neat peaks with an FDR below 0.05 in the distal long arm of AtChr4, which is composed primarily euchromatin and replicates in early S phase (Lee, Pascuzzi, et al 2010).
Short Nascent Strands Treated with Lambda Exonuclease (Exo SNS)

We obtained SNS using lambda exonuclease (Exo SNS). Lambda exonuclease removes all the DNA strands that are not protected with an RNA primer. We obtained three biological replicates to isolate SNS and digested with lambda exonuclease prior to labeling and hybridization to the AtChr4 microarray. We also included technical replicates or dye swaps to account for any dye effect. For the dye swap the 3 bioreps were first labeled with Cy5 and the reference was labeled with Cy3. For the technical replicates, the 3 bioreps were labeled first with Cy3 and the reference was labeled with Cy5. Figure 6 shows a scatter plot matrix comparison of our Exo SNS results with a high correlation between bioreps, ranging from 0.8 between biorep 1 and 3, to 0.9 between biorep 3 and 2. In this case, there is also a high correlation between the technical replicates (Figure 6). Figure 7 shows a comparison of the peaks for the 3 SNS Exo bioreps in a 100-kb region from the 15 to 15.1 Mb region of AtChr4. The Cy5 peaks are shown in red representing the bioreps whereas the Cy3 peaks are shown in green and represent the dye swap or technical reps. It is interesting that, although the agreement between the SNS Exo bioreps are excellent, the correlations are slightly lower than the bioreps in the SNS Neat experiment. However, it is important to note that the agreement shown in the dye swap is also very good in the SNS Exo experiment (Fig 7), in contrast to the lower correlations between dye swap data sets in the SNS Neat experiment.

Since the three bioreps and the three technical replicates had good correlations, the contrast method was used to produce a single dataset. From this dataset, the ratios were determined by dividing the sample by the reference. The ratios were analyzed with the NimbleScan peak finding algorithm and 330 peaks were identified with an FDR below 0.05.
Comparison of the Neat SNS and Exo SNS Profiles

The profiles for both the Neat SNS and the Exo SNS experiments were each highly reproducible. With such high reproducibility, it was important to determine whether the results from the two experimental approaches had any overlap. As can be seen from Figure 8, which shows the same 100-kb region as was shown previously there is a clear overlap between the Neat SNS and Exo SNS data sets.

The extent of overlap of the Neat and Exo SNS over the entire distal long arm of AtChr4 is shown in the Venn diagram (Figure 9). Out of 330 total peaks identified by the SNS Exo, 210 (64%) overlap with the peaks of the SNS Neat profile. Conversely, 217 (30%) of 721 of the SNS Neat peaks overlap with the SNS Exo peaks.

In the Venn diagram (Figure 9) two peaks are scored as overlapping if any bases within one peak are found in the other. Such an approach obviously does not take into consideration the quantitative extent of overlap between two peaks. However, Figure 10 shows that the level of overlap between Neat and Exo SNS is often substantial. Nearly 18% of the Exo SNS peaks and 11% of the Neat SNS have a 100% overlap, and a majority of the peaks scored as overlapping do so by more than 50%. These data show that of the 217 overlapping peaks, the level of Neat and Exo SNS overlap was quite high.

AT Content in Neat and Exo SNS

Our previous analysis of replication timing data showed that initiation regions on AtChr4 tend to be AT-rich (Lee, Pascuzzi, et al. 2010). However, a subsequent study by Costas, et al. 2011) showed that putative origins identified by a BrdU-pulse labeling procedure in A. thaliana cv Landsberg cells have high local GC content (44.5%). Figure 11a...
shows a comparison of the AT content from our Neat and Exo SNS data. The Exo SNS are slightly AT-enriched (64.63%) when compared with the AT content in the AtChr4 DLA (63.66%). Conversely the Neat SNS sequences are slightly AT depleted (60.22%). A comparison of AT-content of the 217 overlapping SNS Exo and SNS Neat peaks shows that their average AT-content is very similar (less than 1% difference) to that of the AtChr4 dla region analyzed (Fig 11b).

**Comparisons with Annotated Genomic Features**

We have also compared the peak locations identified from our Neat-SNS and Exo-SNS experiments with genomic features annotated in TAIR10. Table 1 shows the percentages (bases in a peak that correlate with the genomic feature expressed in a percentage), ratios (bases of the SNS peaks that correlate with the annotated feature divided by the total bases of each feature in chromosome 4) and the p-values that were obtained by the generation of a random dataset permuted 10,000 times (Appendix 1) for some annotations from DLA region analyzed and the corresponding annotation data for Neat and Exo SNS. The annotations compared in Table 1 included: coding sequence (CDS), exon, 5’ UTR (5 prime untranslated region), gene, exon, 3’ UTR (3 prime untranslated region), transposable element (TE), and unannotated. Unannotated features include sequences that are not annotated in TAIR10. Table 1 shows that the Neat SNS signals are reduced in unannotated regions and TEs but generally enriched in gene-containing regions. Conversely Exo SNS are enriched in unannotated regions, but depleted in genes and TEs.

To determine the probabilities that our data randomly aligns with an annotated features, two random data sets were generated, one with the size of neat SNS and another
with the size for Exo SNS. The two random data sets were permuted 100,000 times, which created p values for Exo and neat SNS. Table 2 shows the p values for each feature making it easier to assess the degree of enrichment for various genomic features in regions defined by the Neat and Exo SNS datasets. Regions defined as enriched in the Neat SNS dataset are also enriched in CDS, exon, and genes, and depleted in transposable elements, 5’UTRs and unannotated features. SNS Exo, on the other hand, are enriched in the unannotated features and 3’UTR, and depleted in TEs, genes, CDS, exons and 5’UTR.

Since the regions enriched in the Neat-SNS dataset are enriched for genes, while the Exo-SNS regions are enriched for unannotated features, we examined what genomic features were preferentially associated with regions of peak overlap. The box plots shown in Figure 12 illustrate how the frequencies of the features associated with the overlapping SNS peaks are intermediate between the overall averages for Neat-SNS and Exo SNS, with a slight tendency to be more like the Neat SNS.

**Exo and Neat SNS Compared with Position of Posttranslational Histone Modifications**

We compared our Neat and Exo SNS with some of the reported histone posttranslational modifications (PTMs), H3K56ac, H3K9me2, H3K4 me1/2 and DNA methylation (Tanurdzic et al., 2008). Both the Exo and Neat SNS peaks are depleted for H3K9me2 and H3K56ac, while Neat SNS peaks are enriched for H3K4me1/2 and DNA methylation. These results are consistent with the genomic features described above since the neat SNS are in genes and H3K4me1/2 and DNA methylation are also in genes.
**Exo SNS and Neat SNS Peaks Differ in Position Relative to Genes**

Figure 13a shows a comparison of the distance from the midpoint of either the Neat SNS or the Exo SNS to the closest transcription start site (TSS). Exo SNS peaks are found more frequently upstream of the TSS whereas the SNS Neat signals are more often slightly downstream from the TSS and within the transcribed region. Similarly, Figure 13b shows that the Neat SNS peaks are most often found upstream of the transcriptional termination site (TTS) and within the transcribed regions. Conversely, the Exo SNS enrichment is downstream of the SNS Neat enrichment and tends to be located downstream from the TTS. We cannot decide from this analysis if the SNS is intragenic or intergenic due to the length variation of the genes. The graph shows a 5000-bp region extended to each side of the TSS and TTS. The genes of *Arabidopsis thaliana* have an average of 2500 bp, thus if a SNS is downstream of the TSS, we would not know if it was inside or outside of the gene.

Figure 14a shows the Neat and Exo SNS enrichment relative to a gene. As noted previously the primary regions of Exo SNS enrichment are predominantly upstream of the TSS and downstream of the TTS with some SNS dispersed in the body of the gene slightly enriched towards the 3’ end of the gene body. The Neat SNS enrichment is both downstream from the TSS and upstream from the TTS, and therefore located within the transcribed regions of the gene. Figure 14b shows the same data as Figure 14a, but includes the overlapping SNS.

**SNS and the Replication Timing Profiles**

The goal of the SNS Exo and Neat experiments was to identify origins of DNA replication. When the SNS Neat and Exo results were compared 217 peaks were identified
as putative origins of DNA replication. To validate our results we compared the data from our high-resolution replication timing experiments by Pascuzzi, Lee, et al (in preparation).

From this replication timing data we find that DNA sequences can be replicated early, early-late, early-mid, early-mid-late, intermediate, late, mid and mid-late. The comparison of our data with the replication timing did not show enrichment for any replication time (data not shown). A comparison of our data with initiation, elongation and termination zones of DNA replication showed that the Neat SNS show a slight enrichment for termination zones (data not shown).
DISCUSSION

Putative Origins of Replication Identified from Mapping of Nascent Strands

There is much debate over the identification of origins of DNA replication within the genomes of higher eukaryotes. In Chapter 1, I describe several methods that have been used to identify origins of replication. In the work described in this chapter, we chose to use several methods based on mapping of short nascent strands. For this purpose, SNS can be defined as recently replicated DNA with a single-stranded length longer than Okazaki fragments but still short enough to indicate close proximity to the origin from which they were synthesized. Purification methods include simple size fractionation under denaturing conditions, selecting (for example) DNA fragments with a single-stranded length between 1 and 3 Kb. Alternatively, pulse labeling with BrdU or a similar nucleotide analogue can be used to select for newly replicated DNA prior to the sizing step. It is also possible to exploit the presence of the RNA primer used to initiate replication, which will protect short nascent strands from digestion with lambda exonuclease. Once isolated, putative nascent strands can either be sequenced (Costas et al., 2011) or labeled for hybridization to a microarray, as we have done here.

In preliminary experiments testing the feasibility of using ORC and MCM proteins for ChIP mapping of origins, we found that the large number of ORC2-binding sites along AtChr4 made interpretation extremely difficult (Lee, Franco, et al, unpublished). Although fewer peaks were found following MCM5 ChIP-chip, the lack of reproducibility forced us to use alternative mapping approaches.
We attempted all three SNS methods, but our results for the BrdU-enrichment experiment could not be interpreted because there was a high level of background signal that led to a low correlation between the bioreplicates. However, we found that both the lambda exonuclease (Exo SNS) and the direct size fractionation (Neat SNS) methods produced datasets with excellent agreement between the replicates. These two datasets identified different sets of loci, although a significant subset was identified by both techniques.

We found 721 statistically significant peaks – potential origins - when the Neat SNS procedure was used and 330 with the Exo SNS procedure. In contrast to our previous attempts with other procedures, the reproducibility between replicates of each experiment was very good. In addition, we found that 217 SNS regions were identified by both procedures and the nucleotide overlap was typically greater than 50%. By the total number of peaks identified in the two procedures, we find an average distance of approximately 13 Kb for the Neat SNS, 28 Kb for the Exo SNS and 43 Kb for the overlapping Neat-Exo SNS in the distal long arm of AtChr4. These distances are less that our previous replicon size estimate of 107 Kb (Lee, Pascuzzi, et al. 2010), an earlier estimate of 66 kb (Van’t Hof, et al.1978), or a more recent estimate of 77 kb (Costas, et al. 2011). Although, for the two parental dataset the average distance seem rather small, the overlapping set seems closer to the estimates.

Genomic Analysis of Neat and Exo SNS

There is not a well characterize plant origin of DNA replication, which complicated our interpretation due to the lack of a known positive control for comparison.
Recently Costas et al. (2011) used BrdU ChIP to isolate BrdU-labeled SNS from a sucrose starvation- synchronized *A. thaliana* (Ler) cell culture to identify 1500 putative origins. To validate some of the putative origins, ORC1- and CDC6-binding regions were also isolated by ChIP and found to contain BrdU, providing evidence that the identified regions included newly synthesized DNA.

We know that the replication proteins are largely conserved, which allowed identification of the Arabidopsis and rice homologues (Shultz et al., 2007). However, the DNA sequences with which these conserved proteins interact may vary widely. Some of the common characteristics of animal origins include AT-rich elements (Stanojicic, Lemaitre, Brodolin, Danis, & Mechali, 2008), unmethylated CpG islands, and an overlap with promoters (Delgado, Gomez, Bird, & Antequera, 1998; Sequeira-Mendes et al., 2009), origins formed loops by attaching to the matrix (Courbet et al., 2008; Dijkwel & Hamlin, 1988), and location in nucleosome free regions (Aggarwal & Calvi, 2004; MacAlpine et al., 2010). DNA topology has also been associated with origin activity, with supercoiled regions having a much higher likelihood of binding metazoan ORC (Remus, Beall, & Botchan, 2004; Vashee et al., 2003). Costas et al (2011) reported that *A. thaliana* origins of replication co-map with gene bodies, GC-rich regions, H2A.Z-enriched regions, and regions enriched with H3K4me2, H3K4me3 and H4K5ac (Costas et al., 2011). A recent review from Mechali (2010) indicated similar characteristics in origins of DNA replication in metazoans to those described by Gutierrez’s for plants (Costas et al., 2011). Of particular interest is the observation that Drosophila origins contain repeated GC-rich elements, which have been...
proposed to be a common origin sequence motif shared by all metazoans (Cayrou et al., 2011).

We found that regions identified by the Exo SNS procedure had, on average a slightly higher AT content than the average for AtChr4 DLA, whereas regions identified by the Neat SNS procedure tended to be more GC-rich. However, when the comparison is limited to only regions represented in both the Neat and Exo SNS datasets, the differences in AT-content are insignificant. The AT content in these overlapping regions is also not significantly different from the average of AtChr4 DLA. Therefore, our overlapping SNS peaks seem to differ from those identified by Costas et al, (2011), who concluded that origins have unmethylated GC-rich regions. Our data also does not agree with our previous replication timing results in which the top 25% of the AT-rich probes are more abundant near initiation zones (Lee, Pascuzzi, et al. 2010).

In general, we find that regions identified with the Neat SNS protocol are frequently associated with genes, which are relatively GC-rich compared to the rest of the DLA. However, regions identified using the Exo SNS protocol frequently map to unannotated regions and 3’UTR, both of which tend to be AT-rich. The two data sets must therefore identify largely different populations of sequences. However, we also found that both protocols identify regions of the genome that are depleted in TEs relative to the overall average. These data are consistent with multiple reports showing that regions enriched for TEs tend to be heterochromatic and more likely to replicate in late S phase (Lee et al., 2010; Tanurdzic et al., 2008). Costas, et al (2011) reported that approximately 78% and 10% of their putative origins co-mapped with genes or transposons, respectively. It is important to
note that in our *A. thaliana* cell line, several transposons (TE) families have become active with a concomitant loss of H3K9me2 and an increase in associated 21-nt siRNAs (Tanurdzic et al., 2008). Activation of transposable elements could impact origin-mapping efforts.

Given that some SNS map to gene bodies, we wanted to know more details about their distribution. As expected from our previous results, regions enriched in the Neat SNS protocol tend to be located in the gene body, downstream of the TSS and upstream of the TTS, while sequences enriched in the Exo SNS procedure are predominantly located in non-genic DNA, which agrees with the fact that most of the unannotated features in TAIR10 are in intergenic regions and Exo SNS are also enriched for intergenic regions.

Our previous work showed that H3K56ac is enriched in early replicons and in the initiation zones of both early and late replicons (Lee, Pascuzzi, et al. 2010). However both of our SNS datasets show an unexpected depletion in H3K56ac. The comparison also showed that fragments identified in the Neat SNS protocol are also enriched for DNA methylation, contrary to what is reported for origins in other organisms (Cayrou et al., 2011).

Cayrou, et al (Cayrou et al., 2011), showed that origins from mice frequently occur in the promoter part of the genes, whereas Drosophila origins are spread through the gene body, which is similar to our Neat SNS. Also, Cayrou, et al (2011) showed that Drosophila and mouse origins are enriched in early replication regions. We find no correlation when the Exo SNS peaks are compared with replication timing, while there is a weak preference for mid and late replicating regions when the Neat SNS peaks are compared with replication timing. Neither Exo SNS nor Neat SNS were obviously correlated with the middle point or maximas of the initiation or termination sites (data not shown).
Recent studies of metazoan origins of DNA replication helped to shape a model in which most replication initiates in clusters (or “zones”) of inefficient origins, although a subset of origins have evolved to be well-defined and efficient (Cayrou et al., 2011; Hamlin et al., 2010). If early Arabidopsis origins also fire stochastically (without a determined temporal order) and within a zone of potential origins (Hamlin et al., 2008), mapping discrete sites becomes difficult. Our previous study of replication timing with the same A. thaliana cell line supported the likelihood that replication is stochastic as the replication profiles for early- and mid-S phase were very similar in the euchromatic region of AtChr4 (Lee et al., 2010). In such a case, we could expect to find that SNS mapping would lead to the appearance of having a high level of background that may “map” as a blur of enrichment in a zone, instead of mapping as sharply defined peaks that would be expected if replication always initiated at the same specific site.

Considering all the evidence, we conclude that our SNS data, although highly reproducible as indicated by the high correlation between bioreplicates, cannot unambiguously identify origins of DNA replication. Loci identified in our SNS dataset are on average depleted in H3K56ac, a mark previously been shown to be enriched in initiation zones at the center of replicons (Lee et al., 2010). In addition, the SNS loci do not correlate well with replication timing or with the initiation or termination zones deduced from replication timing data. A possible explanation for our results is that while we may have some putative origins in our datasets, other regions may represent contamination from loci where, for unknown reasons, single-strand breaks are relatively frequent, giving rise to short fragments when DNA is separated on alkaline sucrose gradients. Such breaks may exist in
vivo or, more likely, may be introduced by nuclease activity during the DNA isolation process.

In principle, lambda exonuclease should have degraded all such fragments that were not derived from replication initiation and lacked RNA primers. However, we need to consider the possibility that the use of lambda exonuclease was not efficient enough to digest all DNA fragments lacking primers. Indeed, the difficulty of completely removing such fragments led Mechali’s group to size select first and then expose the selected fragment to two rounds of digestion with a custom-made preparation of lambda exonuclease (Cayrou et al., 2011).

It is possible that the primary sequence is not the major determinant of *A. thaliana* origins of replication, whereas differences in DNA topology may play a major role (Gilbert, 2010). It has been shown that Drosophila ORC has a 30-fold binding affinity for negative supercoiled DNA (Remus et al., 2004). Another study showed that Drosophila ORC correlates with nucleosome free regions (NFR) (MacAlpine et al., 2010; Remus et al., 2004). When an NFR is created by the removal of the nucleosome, this creates negative supercoils, suggesting that ORC binding is not sequence specific but relies in the topology of the DNA.

The lack of information about a well-characterized plant origin of DNA replication greatly limits our confidence in any of the data from our SNS experiments. While the SNS mapping approaches hold promise, being able to find and map a large number of true origins will await the identification of a well-defined, single-copy Arabidopsis origin of DNA replication that can serve as positive control. A positive control will allow us to determine
how well our assays are working in order to produce enough high-quality data to better define origin of replication characteristics.
REFERENCES


## FIGURES AND TABLES

**Table 1. Genomic features and PTMs for Neat and Exo SNS.** Neat and Exo bp= mean the number of bases of a SNS peak that correlate with a genomic feature. Neat and Exo %= is the number of bp but expressed in percentage. Neat and Exo ratio is the % for Neat or Exo SNS divided by the % of chromosome 4. The p-value is the probability of the SNS to be random, a low p-value means the SNS is enriched for the genomic feature.

<table>
<thead>
<tr>
<th>Genomic Feature</th>
<th>Ch4DL (bp)</th>
<th>Ch4DL (%)</th>
<th>Neat (bp)</th>
<th>Neat (%)</th>
<th>Neat Ratio</th>
<th>Neat p-value</th>
<th>Exo (bp)</th>
<th>Exo (%)</th>
<th>Exo Ratio</th>
<th>Exo p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDS</td>
<td>3274025</td>
<td>34.89</td>
<td>386738</td>
<td>52.29</td>
<td>1.499</td>
<td>1.00E-05</td>
<td>114566</td>
<td>27.48</td>
<td>0.7878</td>
<td>0.99999</td>
</tr>
<tr>
<td>Exon</td>
<td>4149382</td>
<td>44.21</td>
<td>446159</td>
<td>60.32</td>
<td>1.364</td>
<td>1.00E-05</td>
<td>151789</td>
<td>36.41</td>
<td>0.8235</td>
<td>0.99999</td>
</tr>
<tr>
<td>5'UTR</td>
<td>322764</td>
<td>3.439</td>
<td>19744</td>
<td>2.67</td>
<td>0.7762</td>
<td>0.98716</td>
<td>11658</td>
<td>2.796</td>
<td>0.8131</td>
<td>0.92004</td>
</tr>
<tr>
<td>Gene</td>
<td>6047070</td>
<td>64.43</td>
<td>603549</td>
<td>81.6</td>
<td>1.266</td>
<td>1.00E-05</td>
<td>238657</td>
<td>57.25</td>
<td>0.8885</td>
<td>0.99647</td>
</tr>
<tr>
<td>TE</td>
<td>547832</td>
<td>5.837</td>
<td>26600</td>
<td>3.596</td>
<td>0.6161</td>
<td>0.99916</td>
<td>11085</td>
<td>2.659</td>
<td>0.4555</td>
<td>0.99967</td>
</tr>
<tr>
<td>Unannot.</td>
<td>2927027</td>
<td>31.19</td>
<td>124991</td>
<td>16.9</td>
<td>0.5419</td>
<td>1</td>
<td>168912</td>
<td>40.52</td>
<td>1.299</td>
<td>1.00E-05</td>
</tr>
<tr>
<td>H3K4me</td>
<td>3219867</td>
<td>34.31</td>
<td>341052</td>
<td>46.11</td>
<td>1.344</td>
<td>1.00E-05</td>
<td>140116</td>
<td>33.61</td>
<td>0.9797</td>
<td>0.48763</td>
</tr>
<tr>
<td>H3K9me2</td>
<td>588091</td>
<td>6.266</td>
<td>39264</td>
<td>5.309</td>
<td>0.8472</td>
<td>0.81964</td>
<td>15224</td>
<td>3.652</td>
<td>0.5828</td>
<td>0.99138</td>
</tr>
<tr>
<td>H3K56ac</td>
<td>3552574</td>
<td>37.85</td>
<td>251935</td>
<td>34.06</td>
<td>0.8999</td>
<td>0.94394</td>
<td>103518</td>
<td>24.83</td>
<td>0.656</td>
<td>1</td>
</tr>
<tr>
<td>DNAme</td>
<td>2746697</td>
<td>29.27</td>
<td>288668</td>
<td>39.03</td>
<td>1.334</td>
<td>1.00E-05</td>
<td>118257</td>
<td>28.37</td>
<td>0.9693</td>
<td>0.54232</td>
</tr>
</tbody>
</table>

CDS= Coding sequence. 5'UTR= five prime untranslated region. TE= transposable elements. Unannot.= Unannotated features on TAIR10.
Figure 1. Flow diagram BrdU SNS. General procedures to obtain BrdU-labeled SNS (BrdU SNS).
Figure 2. Flow diagram Neat SNS. General procedure to obtain SNS selected by size (Neat SNS)
Figure 3. Flow diagram Exo SNS. General steps to obtain SNS using lambda exonuclease
Figure 4. BrdU SNS correlation between microarray bioreplicates. a) Plots for the correlation in between bioreplicates and b) the correlation values between bioreplicates.

<table>
<thead>
<tr>
<th>BrdU SNS</th>
<th>Bioreplicate 1</th>
<th>Bioreplicate 2</th>
<th>Bioreplicate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bioreplicate 1</td>
<td>1</td>
<td>0.288</td>
<td>0.036</td>
</tr>
<tr>
<td>Bioreplicate 2</td>
<td>0.288</td>
<td>1</td>
<td>0.189</td>
</tr>
<tr>
<td>Bioreplicate 3</td>
<td>0.036</td>
<td>0.189</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 5. Neat SNS correlations. a) Plotted correlations between bioreplicates and b) correlation values. The axes are the ratios of sample to reference.

<table>
<thead>
<tr>
<th>Neat SNS</th>
<th>Bioreplicate 1</th>
<th>Bioreplicate 2</th>
<th>Bioreplicate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bioreplicate 1</td>
<td>1</td>
<td>0.901</td>
<td>0.921</td>
</tr>
<tr>
<td>Bioreplicate 2</td>
<td>0.901</td>
<td>1</td>
<td>0.9175</td>
</tr>
<tr>
<td>Bioreplicate 3</td>
<td>0.921</td>
<td>0.9175</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 6. Exo SNS correlations. a) Plotted correlations between bioreplicates and b) correlation values. The axes are ratios of sample to reference.
Figure 7. Comparison of SNS peaks. Comparison of the three bioreplicates (red) and the three technical replicates (green) for a) Exo SNS arrays and b) Neat SNS arrays. Representative window of 100 Kb of *Arabidopsis thaliana* chromosome 4. DS= Dye Swap. BR1= Biorep 1, BR2=Biorep 2. The y axe indicates the ratios sample t reference.
Figure 8. Neat and Exo SNS comparison. First row shows the GC content. Row two shows peaks for Neat SNS, the blue bars show positive peaks. Row three shows Exo SNS data with red bars indicating the positive peaks. Row four shows genes; yellow means the transcript for the gene is present, green means the transcript for the gene is not present and grey is unknown. The direction of the gene is represented by the arrow. Y axe for neat and Exo SNS are ratios sample to reference.
Figure 9. Venn diagram: Agreement between Neat and Exo SNS. 64% peaks of Exo SNS are in common with Neat SNS. The intersection shows a greater number of peaks than the reported for Exo SNS. Sometimes one Neat SNS peak overlaps with two Exo SNS peaks, therefore it is counted twice.
Figure 10. Percentage overlap between Exo and Neat SNS peaks. The red bars represent the Exo SNS peaks that overlap with Neat SNS peaks and the blue bars represent the Neat SNS peaks that overlap with Exo SNS peaks.
Figure 11. **AT content for Neat, Exo and overlapping SNS.** The average AT content for chromosome IV is 63.9%, represented by the dotted line. 

a) **Comparison of Neat–SNS and Exo–SNS AT Content**

b) **Comparison of AT Content for SNS Populations**

The whiskers show little skew of the data and the higher and lower AT content for each dataset. The black dots represent outliers.
Figure 12. Unannotated features for Exo, Neat and overlapping SNS. The vertical dotted line represents the percentage of unannotated features on chromosome IV. The red bars are Exo SNS peaks, showing that the median is close to average content in chromosome 4. The dark blue bars represent Neat SNS peaks. The median is close to zero and has many outliers that are in unannotated features. The orange and light blue bars represent the Exo and Neat SNS overlapping peaks with their median below the average content for chromosome 4.
Figure 13. Transcription Start Site (TSS) and Transcription Termination Site (TTS) for Neat and Exo SNS. a) Exo SNS are mainly upstream of the TSS, while Neat SNS are predominantly downstream of the TSS. b) Neat SNS are mainly upstream the TTS and Exo SNS are around the TTS. The y axes are the SNS peaks percentage distribution, and in the X axes indicates the TTS or TSS of a gene with a 5000-bp distance upstream and downstream.
Figure 14. **SNS position relative to a proximal gene.** The SNS was measured to their most proximal gene. a) Exo and Neat SNS. Exo is predominantly located upstream and downstream of the gene with some SNS in the gene body. Neat SNS are located in the gene body. b) Neat/Exo and Exo/Neat overlapping SNS, compared with the parental SNS.
APPENDIX
Appendix A. Random permuted data and observed values of enrichment relative to AtChr4 distal long arm. Two random datasets were generated, one for the Exo SNS and another for the Neat SNS, and permutated 100000 times to estimate p-values. Each graph represents the data for a specific genome feature (Exon, genes, transposable elements, etc). The continuous lines represents the permutations while the dashed vertical lines represent the observed values for Neat SNS (blue) and Exo SNS (red). Non-random data will show outside of the curve of the continuous line.
Permutation Results for Exon

Counts

Enrichment Relative to Chr4–DL

Random Neat–SNS
Actual Neat–SNS
Random Exo–SNS
Actual Exo–SNS
Permutation Results for Five Prime UTR

Counts

Enrichment Relative to Chr4–DL

Random Neat–SNS

Actual Neat–SNS

Random Exo–SNS

Actual Exo–SNS
Permutation Results for Gene

Counts

Enrichment Relative to Chr4–DL

Random Neat–SNS
Actual Neat–SNS
Random Exo–SNS
Actual Exo–SNS
Permutation Results for Transposable Element

Counts

Enrichment Relative to Chr4–DL

0 0.5 1.0 1.5 2.0

Random Neat–SNS
Actual Neat–SNS
Random Exo–SNS
Actual Exo–SNS
Permutation Results for Unannotated

Counts

Enrichment Relative to Chr4–DL

Random Neat–SNS

Actual Neat–SNS

Random Exo–SNS

Actual Exo–SNS
Permutation Results for DNAme

![Graph showing enrichment relative to Chr4-DL]
Permutation Results for H3K4me

Counts

Enrichment Relative to Chr4–DL

Random Neat–SNS

Actual Neat–SNS

Random Exo–SNS

Actual Exo–SNS
Permutation Results for H3K9me2

Counts

Enrichment Relative to Chr4–DL

Random Neat–SNS
Actual Neat–SNS
Random Exo–SNS
Actual Exo–SNS
Permutation Results for H3K56ac

Enrichment Relative to Chr4

Counts

Random Neat–SNS
Actual Neat–SNS
Random Exo–SNS
Actual Exo–SNS

Enrichment Relative to Chr4–DL
CHAPTER III

Changes in Histone Modification during G1, S and G2/ early M
INTRODUCTION

The eukaryotic genome must be packaged into a nucleus in a manner that is highly organized to facilitate the critical nuclear processes necessary for survival of the cell. The DNA is packaged by interaction with histone proteins, which allows the entire genome to fit within the nucleus. Genome packaging allows dynamic interactions to occur by controlling DNA accessibility to regulatory factors by modulating nucleosome position and local chromatin structure. Chromatin structure and DNA accessibility are controlled by chromatin remodeling, a dynamic process that leads to specific histone and DNA modifications (O. Bell et al., 2011).

To understand the how chromatin structure is controlled it is necessary to know the basic levels of chromatin structure. The diameter of the double helix DNA ranges from 2 to 2.3nm. The DNA is wrapped twice around a histone octamer (nucleosome), which has a diameter of 11 nm. The nucleosome is composed of 2 molecules of each core histone (H2A, H2B, H3 and H4). The presence of the histone H1 links the nucleosomes, further packaging the DNA and resulting in the formation of the 30-nm filament, although there is still debate about the existence of 30 nm fibers and other models exist for this level of organization. Higher-order packaging beyond the 30-fiber is still largely unknown (Fussner et al., 2011; Maeshima et al., 2010). Current models predict that 30-nm chromatin fibers form ~100-nm loops with the base of the loop attached to a matrix or scaffold (Flors & Earnshaw, 2011). Finally, series of loops coil make a chromatid and two chromatides makes a chromosome (Bloom & Joglekar, 2010).
The DNA is not always tightly packed into chromosomes, contrary to what one might think. It is only packed when its needs to be equally divided during mitosis, so daughter cells have the same number of chromosomes and the same genetic information (Bloom & Joglekar, 2010). Chromatin structure is re-organized as the cell passes through interphase (G1, S and G2/ early M) and begins mitosis. The processes of replication and transcription require the DNA to be accessible to allow binding by the replication machinery, transcription factors, RNA polymerases, and other factors. Euchromatin is accessible chromatin, whereas partially inaccessible, or totally inaccessible chromatin is either facultative or constitutive heterochromatin, respectively. The process of chromatin remodeling controls the type of histone modification and determines whether the chromatin is accessible or inaccessible.

Each core histone in the nucleosome can be modified (remodeled). Initial studies of histone modification (Tamaru et al., 2003) showed that various lysines within the histone H3 tail could be methylated or acetylated, the serines phosphorylated and the arginines methylated. Histone modifications can have very different effects on chromatin accessibility with modifications such as H3K9me1, H3K9me2, or H3K9me3 causing changes that have important effects on accessibility. In Arabidopsis thaliana, H3K9me1 and H3K9me2 are associated with chromocenters (Fransz & de Jong, 2011), which are typical epigenetic marks for inaccessible (inactive) chromatin, whereas H3K9me3 is associated with accessible (active) chromatin (Berger et al., 2011) reviewed in (C. Liu et al., 2010).

The majority of studies on histone post-translational modifications (PTM) address how the modification alters transcriptional activity, or gene expression (RNA levels). However, the importance of histone PTM during chromatin remodeling is also an important
process that controls DNA replication. The tremendous number and combinations of histone PTMs that are possible create a vast array of potential conformations that chromatin may take (Garcia et al., 2007).

I will summarize briefly what is known about the histone PTMs that are thought to play roles in DNA replication and how they are modulated during DNA synthesis. It has been reported that acetylation of H3K18, H4K16, H4K5, H4K8, and H4K12 are related to new histone deposition during DNA synthesis in plants (Chen & Tian, 2007; Fuchs, Demidov, Houben, & Schubert, 2006); and reviewed in (Sanchez Mde, Caro, Desvoyes, Ramirez-Parra, & Gutierrez, 2008), humans and drosophila (Aline V. Probst et al., 2009). Schwaiger, Stadler, et al (2009) reported that H4K16ac is associated with early replication in drosophila. Costas, et al (2011) found an enrichment of H4K5ac at putative origins of replication in the A. thaliana genome (Costas et al., 2011).

Our previous study (Lee et al., 2010) showed H3K56ac enrichment in early replicons and at expressed genes from chromosome 4 of A. thaliana (Lee et al., 2010). However, the molecular mechanism that H3K56ac plays during DNA replication is largely unknown for higher eukaryotes. Recent studies of S. cerevisiae described a role for H3K56ac in stabilizing replication forks (Clemente-Ruiz, Gonzalez-Prieto, & Prado, 2011; Tsubota et al., 2007) and promoting new histone deposition by the chromatin assembly factors CAF1 and histone chaperone Asf1 (Li et al., 2008). During replication Asf1 binds to newly synthesized H3/H4 dimers (Tyler et al., 1999) and enhances the acetylation of the H3K56, which is catalyzed by the acetyltransferase Rtt109 (Driscoll, Hudson, & Jackson, 2007). H3K56ac then enhances the H3 binding affinity for CAF1 and Rtt106, and CAF1 binding to PCNA at the replication
fork, resulting in deposition of new histones (Clemente-Ruiz et al., 2011; Li et al., 2008). An *A. thaliana caf-1* mutant has a greatly increased frequency of somatic homologous recombination (~40-fold) and T-DNA integration compared to the wild-type (Endo et al., 2006). *A. thaliana* has two *ASF1* genes (*AtASF1A* and *AtASF1B*) and the *Atasf1ab* double mutant has reduced cell number, S-phase delay/arrest, and reduced ploidy levels (Zhu et al., 2011). The *S. cerevisiae* double mutant *asf1D/rtt109* cannot acetylate H3K56 and has an increased frequency of homologous recombination and gross chromosomal rearrangements (Driscoll et al., 2007; Han et al., 2007; Prado, Cortes-Ledesma, & Aguilera, 2004).

H3K27me3 has been described as a mark for the polycomb group (PcG), which dynamically controls developmental genes by silencing (Bouyer et al., 2011; Muller & Verrijzer, 2009). A recent study in Arabidopsis showed that H3K27me3 dynamically regulates developmental genes during cell differentiation and is found in differentiated and undifferentiated cells (Lafos et al., 2011). This repressive mark is found in euchromatin (Roudier et al., 2011). While the role of H3K27me3 during replication is unknown, Hansen, et al (2008) proposed a model for how the H3K27me3 mark is propagated through S phase in human cells. They proposed that H3K27me3 recruits the Polycomb Repressive Complex 2 (PRC2) to mark a specific site during DNA replication. The PRC2, which is retained at the site during replication, can then re-establish the H3K27me3 mark on the daughter strand, thus allowing chromatin structure and transcriptional programs to be transmitted between generations (Hansen et al., 2008).

H3K9me2 is a repressive mark found primarily in heterochromatin and is associated with transposable elements (TEs) and methylated repeat sequences (Tanurdzic et al., 2008).
Costas, et al (2011) found that H3K9me2 is depleted at putative origins of DNA replication of *A. thaliana*. Immunolocalization studies showed that H3K9me2 co-localizes to chromocenters along with H3K9me1, H3K27me1, H3K27me2 and H4K20me1 (Fransz, ten Hoopen, & Tessadori, 2006).

H3S10p is a mitotic mark described as a mark for chromosome condensation and segregation in plants (Granot, Sikron-Persi, Li, & Grafi, 2009; Houben et al., 2007) and mammals (Hendzel et al., 1997). The role of H3S10p during DNA replication has been extensively studied in *S. pombe* centromeres. During M phase H3S10 phosphorylation is correlated with decreases in centromeric heterochromatin levels by preventing the binding of Swi6 (HP1 in mammals), a chromodomain protein, to H3K9me. During S phase, H3S10p is removed and replaced by H3K9me, which is again bound by Swi6 to re-establish the heterochromatic marks in the newly synthesized DNA (reviewed in (Beisel & Paro, 2011). H3S10 phosphorylation is particularly important during the replication of centromeric DNA as the loss of Swi6 results in the transcription of centromeric repeats in early S phase (Beisel & Paro, 2011; Kloc, Zaratiegui, Nora, & Martienssen, 2008).

H3S10p is present in both interphase and mitotic nuclei of mouse embryonic cell (EC) lines and enriched on pericentric heterochromatin during late S and G2 phases of the cell cycle (Fazzio & Panning, 2010; Hendzel et al., 1997). However, when condensin, a complex responsible for the structural maintenance of chromosomes, is knocked down, ES nuclei are depleted of H3S10p in interphase nuclei, whereas H3S10p levels are maintained in mitotic chromosomes (Fazzio & Panning, 2010). When other epigenetic marks in the condensin knockdown nuclei were examined, only H3S10P, H3K9me3, and 5meC had
altered localization patterns. The *S. pombe* and mouse embryo indicated that H3S10p localization is antagonistic to the heterochromatic mark H3K9me3 (Beisel & Paro, 2011; Fazzio & Panning, 2010). In Drosophila, H3S10p is important for early transcription elongation by binding to 14-3-3 proteins (Karam, Kellner, Takenaka, Clemons, & Corces, 2010) along with occupying promoters to prevent H3K9 methylation keeping them poised for activation (T. K. Kelly et al., 2010; Nowak & Corces, 2004). While studies of H3S10p are limited in plants, it appears to have a similar role as in other eukaryotes (Houben et al., 2007; Houben, Demidov, Rutten, & Scheidtmann, 2005; Houben et al., 1999).

To date, very little information is available on the histone modifications that occur during DNA synthesis. In the present study, we used FACS to sort the nuclei of *A. thaliana* suspension cells into G1, S, and G2/early M populations. The sorted nuclei were compared using immunoblot analysis to determine whether a change in the amount H3K56ac, H3S10p, and H4K16ac relative to total H3 occur during G1, S, and G2/early M phases of the plant cell cycle. We found no clear change in the levels of the selected histone PTMs.

While no change in the levels of H3K56ac, H3S10p, and H4K16ac were detected, these results do not rule out the possibility that the location of the modified histones change in the nucleus. To address possible changes in the localization of the modified histones, we used confocal microscopy and immunolocalization. Localization patterns of H4K5ac, H4K16ac, H3K56ac, H3K9me2, H3K27me3 and H3S10p were compared. We found no clear change in the localization of H3K56ac or H4K16ac. However, while the amount of H3S10p remains unchanged in G1, S, and G2/early M nuclei, the G2/early M pattern showed increased H3S10p at the periphery and clear localization within the nucleolus. H3K27me3
has a punctuate pattern, but a more detailed analysis using higher resolution would be useful in order to understand the significance of the pattern we observe.
Cell Culture

The *Arabidopsis thaliana* Col-0 cell line was maintained as previously described (Lee, Pascuzzi et al. 2010). To produce a ‘7-d split culture’, 25 mL of fresh medium was mixed with 25 mL of a 7-d Arabidopsis culture. The 7-d split culture was grown for 17 h and the cells were harvested as previously described (Lee, Pascuzzi et al. 2010). The cells were washed three times in 1X phosphate buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄·2 H₂O, 2 mM KH₂PO₄, pH 7.4). Excess buffer was removed using a filter aspirator, snap frozen in liquid nitrogen, and stored at -80°C until use. One biological replicate comprised 4 flasks of the ‘7-d split culture’.

Nuclei Isolation

The frozen cells were gently ground in liquid nitrogen by mortar and pestle. The cell powder pellet was immediately resuspended into 50 mL of cold nuclei isolation buffer (NIB) (500 mM Hexylene glycol, 20 mM HEPES pH 7.4, 20 mM KCl, 0.5 mM EDTA, 0.5% Triton, 1 mM DTT, 1 mM MgCl, 1 mL Protease inhibitor cocktail (Sigma P-9599), 10 mM Sodium butyrate, 50 mM Nicotinamide, 1 mM Orthovanadate and 20 mM NaF). The suspension was stirred briefly, passed through Miracloth and then a 3-tiered nylon mesh (100, 50, and 30 µm) and the filtrate was centrifuged at 239 x g for 10 minutes at 4°C. The pellet was resuspended in 6 mL of the same modified NIB buffer. Two µL of DAPI (2 mg mL⁻¹) were added after the nuclei were passed through a 20 µm filter prior to sorting.
Flow Sorting and Protein Extraction

The nuclei were checked by microscopy for yield and purity before sorting by DNA content with an InFlux cell sorter (Becton Dickinson, Biosciences) equipped with a 355-nm UV laser and a 488-nm sapphire laser. The sheath fluid was 1 x STE (10 mM Tris pH7.5, 10 mM EDTA and 100 mM NaCl). Sorted nuclei were recovered in 5 mL tubes containing 2 mL STE with 2 tablets of proteinase inhibitor (Roche Complete tablets) per 10 mL and 50 mM Nicotinamide.

A total of 12.5 X 10^6 nuclei were sorted, which typically yielded 6 X 10^6 G1, 4 X 10^6 G2, and 2.5 X 10^6 S phase nuclei per FACS run. The sorted nuclei were centrifuged at 4,000 x g for 10 minutes at 4°C. To extract the nuclear protein, the nuclei were incubated in ice for 30 min in the amount 2X Laemmli buffer (Laemmli, 1970) necessary for a final concentration of 20,000 nuclei µL^{-1}, and sonicated twice at 25% amplitude for 10 seconds and placed on ice for 45 seconds. A Sonicator Vibra cell VCX 130 with CV18 tip Model (Sonic and Material, Inc, Newtown. CT) was used for our experiments. The resulting samples were stored at -20°C in ~50 µL aliquots.

Western Blots and Statistical Analysis

Mini 15% acrylamide SDS-page 1.5 mm thick with a 4% stacking gel was used to analyze the content of the various modified histone from the total nuclear proteins preparation. The samples were electrophoresed for 40 minutes at 185 V in 1x running buffer from Bio-Rad (10X Tris/Glycine/SDS buffer, cat# 161-0732). The proteins were electrophoretically transferred from the gel using 1x transfer buffer (192mM Glycine, 50mM Tris, 20% Methanol and water). PVDF (poly-vinylidene fluoride) membranes were pre-
wetted in methanol and equilibrated in 1x transfer buffer before transfer. After transfer, the membranes were blocked in 1x TBS (Tris buffer saline: 50 mM Tris HCl, 150 mM NaCl, pH 7.4) containing 3% skim milk for 1 h. For immunological detection, membranes were incubated with primary antibody diluted in blocking buffer (blocking buffer for fluorescent Western blotting, cat # MB-070, Rockland, Inc.) with gentle rocking overnight at 4°C. Membranes were then rinsed three times with distilled water and incubated for 2h at 4°C in secondary antibody diluted 1:5000 in blocking buffer. After the second incubation, membranes were washed three times for 5 minutes each with 1x TBST (TBS plus 0.1% tween 20). The primary antibody concentration varied depending on the histone modification being analyzed (Table 1). Membranes were scanned in the Odyssey scanner (Li-cor) at a wavelength of either 700 nm or 800 nm depending on the secondary antibody (Table 1). A paired T-test for two bioreps, in which each biorep included 4 technical replicates (Ott, 2001).

**Cell Fixation for Immunolocalization of Histone Modifications**

*A. thaliana* suspension cells were harvested after the “7 d split” as described above by centrifuging for 5 minutes at 250 x g at room temperature. The supernatant was discarded, and the cell pellet re-suspended into 40 mL fixation buffer (1% v/v para-formaldehyde in 1X PBS) and incubated for 15 minutes, mixing every 3 minutes. The fixation was stopped with 2.5 mL of 2M Glycine, and the cells were washed 3 times with 1X PBS. The buffer was removed by vacuum filtration and the cells were snap-frozen in liquid nitrogen prior to storage at -80 °C.
Nuclei Isolation for Immunolocalization of Histone Modifications

Five mL of cold NLB (Nuclei Lysis Buffer; 50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS plus 1 tablet Protease inhibitor from Roche Complete) was added to the frozen fixed cells to dislodge the pellet from the tube. The frozen pellet was poured into a commercial blender (Cuisinart “smart power duet”) containing 65 mL of pre-cooled NLB and “chopped” according to a protocol designed by Dr. Tae-Jin Lee (Lee et al., 2010). After blending on the “chop” setting for one minute, the mixture was allowed to settle and thaw for 5 minutes with a brief “chop” at approximately 2.5 min. The mixture was then passed through 2 layers of Miracloth and 3 layers of nylon mesh (100, 50, and 30 µm). The filtered solution was centrifuged at 250 x g for 5 minutes at 4°C and the nuclear pellet was resuspended in 0.5 mL of NLB. The suspension was passed through a 20-µm filter and DAPI (2 µg/mL) was added at least 5 minutes before sorting.

Nuclei Sorting for Immunolocalization of Histone Modifications

Nuclei were sorted according to DNA content for G1, S and G2 phase directly onto pre-cleaned positively charged slides (Fisher cat# 22-230-900). Each slide, with 4,000 sorted nuclei was then covered to dry and stored at -80°C until further use.

Immunolocalization of Histone Modifications

For immunolabeling, each sample was first blocked with 250 µL of normal goat serum diluted 1:10 in 1X PBS for one hour at room temperature. The primary antibody, diluted at a concentration determined by previous experiments (Table 3) was then added, and the sample was incubated either for 2 hours at room temperature or overnight at 4°C. The slides were washed three times for five minutes with 1X PBS. The secondary antibody,
diluted in 1X PBS, was then added, the sample was incubated in the dark for 2-4 hours at room temperature and washed four times with 1X PBS as described above. DAPI (0.5 µg mL⁻¹) was included in the final 1X PBS wash. ProLong Gold antifade reagent (Invitrogen) was used as a mounting medium following the manufacturer’s protocol. The slides were sealed with nail polish and stored in the dark at room temperature.

**Confocal Microscopy**

Images were obtained with a Zeiss LSM 710 confocal microscope with a 40X water immersion objective and processed with Zen 2009 light edition software. Most of the modified histones yielded strong fluorescence signals, with the exception of H3K9me2 and H3S10p, which were very weak and required a longer exposure to the laser, resulting in bleaching of the fluorescence. Initially Vectashield (Vector Labs) was used to reduce bleaching, but the resulting blue hue interfered with fluorescence detection. ProLong Gold (Invitrogen) mounting medium was then used, which reduced bleaching without interfering with detection of the fluorescent antibodies. Unfortunately, not all of the histone PTM samples were mounted with ProLong Gold, and some images were of lower quality. However, despite the reduced quality, it was still possible to determine the overall nuclear localization patterns. Our primary goal was to determine whether localization patterns of the modified histones changed during the cell cycle, so laser intensity was varied depending upon the level of modified histone fluorescence signal. For example, when high laser intensity was used the fluorescence signal of H3-CT faded rapidly. Therefore, during image processing, the Zen 2009 light edition software was used to enhance the H3 CT signal, to allow visualization for comparison with both DAPI and the specific modified histone.
Notes

Different methods were tried to isolate histones. The high salt extraction protocol described by Sandra Hakes’ laboratory ([http://www.molekularbiologie.abi.med.uni-muenchen.de/ueber_uns/hake/index.html](http://www.molekularbiologie.abi.med.uni-muenchen.de/ueber_uns/hake/index.html)) was difficult to apply to our situation because it required a large amount of starting material.

It is better to use PVDF membranes due to the small weight of the histones. It is also best to avoid thawing and freezing of the samples. Overnight incubation with the primary antibody overnight gave better results than 1 or 2 hours at room temperature.

To obtain immunoblots for publication, NuPAGE 4-12% Bis-Tris acrylamide gels from Invitrogen were used with the recommended buffer and protocol. Transfer conditions and blotting conditions remained the same.
RESULTS

Quantifying Modified Histones in G1, S, and G2 Nuclei

We quantified the levels of total nuclear H3K56ac, H3S10p, and H4K16ac, which have been shown to be either associated with DNA replication or increase chromatin accessibility during early replication. Lee et al (Lee, Pascuzzi, et al, 2010) showed that H3K56ac was associated with early replicating regions of A. thaliana chromosome 4, whereas H3S10p is a marker for mitosis (M phase) (Fischle et al., 2005) H4K16ac is associated with accessible chromatin during early DNA replication in Drosophila (O. Bell et al., 2010).

Nuclei were isolated from 7-day split Arabidopsis suspension culture cells (Lee, Pascuzzi, et al. 2010) and sorted into G1, S, and G2/early M populations. Each fraction was solubilized in SDS-PAGE loading buffer, appropriate aliquots were loaded onto polyacrylamide gels for immunoblot analysis, and the signals were detected and quantified using an Odyssey scanner. The abundance of each modified histone was compared to the level of total histone H3 unmodified C-terminus, detected with an antibody to the unmodified C-terminus (Table 2) to normalize for variations in extraction efficiency and loading. Representative results for each of the modified histones are shown in Figure 1 and Table 2.

We were unable to detect any differences in the nuclear abundance of H3K56ac, H3S10p, or H4K16ac in S phase as compared to G1 or G2/M nuclei. This conclusion is based on two biological replicates with four technical replicates for each histone PTM. Table
2 shows that paired t test analysis detected no significant differences in the levels of H3K56ac, H3S10p, or H4K16ac levels between G1, S, or G2/M phase (Table 2).

While some minor changes were seen in H3S10p levels during the cell cycle, the signal was very low which gave a wide range of values for each technical replicate. The variation in the H3S10p levels is likely due to the lower signal, which made the quantification more difficult as the signal to noise. Decreases. For example, the t test for biorep one indicated no significant difference for H3S10p, whereas t test analysis of biorep 2 shows significant differences between G2-G1 and S-G1 phase (Table 2). While some signal increase was found during S phase, that was retained in G2/early M, there is a strong blot-to-blot variation for H3S10p. Therefore, if there is a difference in H3S10p levels, it is obscured due to the variation.

**Localization of the Posttranslational Marks During the Interphase**

Our immunoblot analyses revealed no detectable changes in the total abundance of H3K56ac, H3S10p, or H4K16ac during G1, S, and G2/early M phases, but does not address whether changes in the subnuclear localization had occurred. Therefore, we then examined the localization patterns for these modified histones as a function of cell cycle stage. Fixed nuclei from G1, S and G2/early M phases were sorted on to charged slides for immunolocalization by confocal microscopy. Each slide included 4,000 G1, 4,000 S and 4,000 G2/early M phases nuclei, was immunolabeled with a specific antibody, corresponding to either: H4K16ac, H3K56ac, H3K9me2, H3S10p or H4K5ac. At the same time each slide was immunolabeled with the anti-H3CT in order to compare the localization of the histone
PTM relative to the total H3-CT, and stained with DAPI to determine the location of the nuclear DNA.

G2 nuclei are larger than G1 nuclei, the S phase nuclei range in size from what is found in G1 to the sizes seen for G2. This observation is consistent with DNA replication occurring during S phase to G2, and confirms that the sorting was successful. For comparison of histone PTM, approximately 300 nuclei were examined for each treatment.

The different histone PTMs visualized included H3K9me2, H3K27me3, H3K56ac, H4K5ac, H4K16ac and H3S10p (Figure 2). Immunoblots H3K9me2 had no signal (data not shown). Because H3K9me2 is an important mark for heterochromatin, it was included in our localization study. H4K5ac has been shown to be important during the de novo deposition of histones (Fuchs et al., 2006; Aline V. Probst et al., 2009; Sirbu et al., 2011).

H4K5ac is easily visualized and is scattered evenly through the nucleoplasm but absent in the nucleolus (Fig 2A). H4K5ac localization remains essentially the same in the G1, S and G2/ early M nuclei. Similarly, H4K16ac is also easily visualized, evenly distributed throughout the nucleoplasm, absent in the nucleolus (Fig 2B), and shows no discernable difference between the G1, S and G2/early M nuclei. Figure 2C shows that although the H3K56ac signal is weaker than either H4K16ac or H4K5ac, the localization patterns are similar, with an even distribution throughout the nucleoplasm and absence from the nucleolus. Since the slides for H3K56ac localization were not mounted with the ProLong Gold, the fluorescence signal faded quickly during laser scanning, and the images are not as clear as those for H4K5ac and H4K16ac. Regardless, it is still apparent that not differences were detected in the H3K56ac localization patterns (Figure 2C).
Figure 2D shows that the H3K9me2 localization pattern is different than the patterns for the acetylated histones H4K5, H4K16 and H3K56. The H3K9me2 fluorescence signal is very weak and does not appear to co-localize with the H3-CT, but the bleaching of H3-CT makes it difficult to make a clear conclusion. However, it is clear that H3K9me2 is unevenly distributed and exhibits a punctate pattern, with some increased concentration at the nucleolar periphery and within the nucleolus. However, comparison of the H3K9me2 nuclear localization patterns for G1, S, and G2/ early M nuclei shows no clear changes in pattern.

H3K27me3 is localized throughout the nucleoplasm in G1, S, and G2/ early M nuclei and is absent from the nucleolus (Fig 2E). Our images also show large foci with higher levels of H3K27me3 fluorescence that are most obvious in G2/ early M.

As expected from the immunoblot analysis, the H3S10p signal (Fig 2F) is very weak, but does appear to increase in G2/ early M relative to G1 or S phase. A punctate pattern is also seen with the H3S10p, with some signal in the nucleolus in the G2/ early M images. We could not confirm if H3S10p is inside the nucleolus because the laser intensity required to see the fluorescence signal caused rapid fading, although the Z stack images suggest that some of the localization is truly nucleolar. While the H3S10p distribution appears throughout the nucleus in G1 and S, the nuclei in G2/ early M clearly show that H3S10p form foci both within the nucleolus and at the periphery. It is also possible that foci located within the nucleolus in G2/early M phase, may be the events of early condensation of the Arabidopsis chromosomes.
DISCUSSION

H4K5ac, H4K16ac, H3K56ac, and H3S10p Levels Remain Unchanged During the Cell Cycle

How parental histones are reassembled into nucleosomes following DNA replication is poorly understood. It is known that the DNA replication complex must coordinate with histone remodeling factors to gain access to the chromatin template both before and during active replication (Aggarwal & Calvi, 2004). One of the best-known histone remodeling pathways that is known to increase chromatin accessibility is histone acetylation. During the early studies of histone remodeling, acetylation was associated with transcriptional activity. More recently histone acetylation has also been shown to be associated with DNA replication (Danis et al., 2004; J. Liu, McConnell, Dixon, & Calvi, 2012; Vogelauer, Rubbi, Lucas, Brewer, & Grunstein, 2002).

In the legume *Vicia faba*, the relative abundances of H4 acetylation is modulated during the plant cell cycle as the presence of acetylated H4K5 in nucleolar organizer regions increased during M, but decreased during S as the nucleolus appeared to “transiently” lose nucleosomes (Jasencakova, Meister, Walter, Turner, & Schubert, 2000). In Arabidopsis, the H416ac localization in root tip nuclei is similar to *Vicia faba*, and is the only acetylated histone modification present in the nucleolus (Jasencakova et al., 2003). Arabidopsis H4K16ac is also dispersed evenly and abundantly throughout the nucleoplasm (Jasencakova, Meister, & Schubert, 2001; Jasencakova et al., 2000; Jasencakova et al., 2003). H4K16ac increases in both euchromatic and heterochromatic regions during replication (Fuchs et al.,
and at early-replicating regions of *Vicia*, Arabidopsis (Jasencakova et al., 2000; Jasencakova et al., 2003) and Drosophila (Schwaiger et al., 2009). Our group (Lee et al., 2010) recently reported enrichment of acetylated H3K56 in early replicons and initiation zones of both early and late replicons in *Arabidopsis thaliana*.

In the present study, we found no variation in the total nuclear abundance of H3K56ac or H4K16ac during the cell cycle. Although at first glance, this result may seem to contradict our observation of H3K56ac enrichment in early replicating regions, it is more likely that the total amount H3K56ac (or of H4K16ac, presumably) is large in relation to the amount in replicating regions, so that changes associated with replication are limited to a small percentage of the total pool. Genomic locations enriched in H3K56ac by chromatin immunoprecipitation using our high resolution NimbleGen microarray indicates that localization of this mark is largely the same in S phase as in either G1 or G2/M. At a few specific locations, however, levels of H3K56ac are specifically increased during S phase (Pascuzzi, Lee, in preparation). Subtle differences of this type would be undetectable in our immunoblot analysis of whole nuclei.

It should be noted that the same type of assay is possible for *Saccharomyces cerevisiae*. Matsumoto using *Saccharomyces cerevisiae* synchronized by sucrose starvation found a transient increase in H3K56ac levels 45-75 minutes after G1 release (Masumoto, Hawke, Kobayashi, & Verreault, 2005). In humans and *Saccharomyces cerevisiae*, newly synthesized H3 and H4 are acetylated by a histone acetylase (HAT) and form tetramers in the cytoplasm. Chaperones then facilitate the import of the H3–H4 tetramers into the nucleus.
where ASF1 and CAF then mediate their assembly onto the newly replicated DNA (Allshire & Karpen, 2008).

Our data suggest that the pools of free acetylated H4K16ac and H3K56ac remain high throughout the cell cycle, masking variations in chromatin bound acetylated histones, and/or that variations in localization may occur in the absence of detectable changes in total abundance of chromatin-bound acetylated histones.

Phosphorylation of H3S10p has been well characterized and is conserved among higher eukaryotes during mitosis. H3S10 is phosphorylated during G2 with the signal peaking at metaphase and then decreasing during the cell cycle. H3S10 plays an important role in condensation and segregation of chromosomes (Nowak & Corces, 2004). Due to its role during mitosis, it is typically used as a marker for mitotic cells (Wei, Yu, Bowen, Gorovsky, & Allis, 1999). We found that the nuclear pool of H3S10p is very low in G1 and S nuclei and thus difficult to quantify, but we saw slight increase in G2/ early M nuclei. Our data along with the fact that H3S10p is a mark for M phase suggests that any increase we observe may be due to some M phase nuclei in our G2 nuclei samples, but the high variability we find in the G2/ early M H3S10p levels makes this conclusion speculative.

**Subnuclear Localization of Histone Modifications During the Cell Cycle**

Although we found little or no change in the abundance of total nuclear H4K5ac, H3K56, and H3S10p during the cell cycle, it remains possible that DNA replication is accompanied by significant changes in the sub-nuclear localization patterns for one or more of these modified histones. To investigate this possibility, sorted nuclei were labeled with a
primary antibody against the modified histone and a secondary antibody conjugated to a fluorophore, and examined by confocal microscopy.

The sub-nuclear distribution of each histone acetylation mark (H4K5ac, H4K12ac, and H3K56ac) remained unchanged in G1, S, and G2/Early M nuclei. The pattern was spread homogenously through the cell with no signal in the nucleolus. This is not surprising as histone acetylation is generally associated with accessible chromatin, which is the predominant form during DNA replication consistent with the observed dispersed signal throughout the nucleus.

The images obtained with the confocal microscope did not reveal clear global differences in the localization patterns of H4K5ac, H4K16ac, and H3K56ac between G1, S and G2/early M phases, indicating that any changes in localization were below our limit of detection. Our H4K16ac localization results also seem to differ from the previous results of Jasencakova et al (2003), who found that H4K16ac was also localized within the nucleolus of Arabidopsis root tip nuclei. While further analysis is necessary to make a conclusion, we used nuclei from suspension cells, which may be quite different from the nuclei of differentiated tissue (Jasencakova et al, 2003).

**H3K27me3 and H3K9me2 Variation During the Cell Cycle**

In higher eukaryotes the repressive H3K27me3 mark is typically found embedded in euchromatic regions (Lafos et al., 2011; Peters et al., 2003). This observation is consistent with our finding that H3K27me3 exhibits a punctate pattern that appears to be independent of the brightest DAPI signals, suggesting that H3K27me3 is indeed localized in regions where the DNA is less condensed. The pattern seems to be strongest in G2/early M nuclei, but the
difference is subtle. H3K27me3 is associated with Polycomb proteins (PcG) in both plants and animals. Many PcG proteins are repressive and associated with development (Desvoyes, Sanchez, Ramirez-Parra, & Gutierrez, 2010; Lafos et al., 2011) but may fill a different role in our cell culture system.

H3K9me2, which is associated with silenced transposable elements (Tanurdzic et al., 2008), was of low-abundance and difficult to visualize in our suspension cell line. However, our data are consistent with the study by Tanurdzic et al (2008), who used ChIP-CHIP to analyze AtChr4 regions bound by H3K9me2 in the same Arabidopsis cell line. They found that specific families of TEs that are typically hypomethylated had lost H3K9me2 in the cell line (Tanurdzic et al., 2008).

Another finding from our study is the presence of H3K9me2 in the nucleolus of suspension cell nuclei. Costa-Nunes recently showed that decondensation of rDNA foci in A. thaliana is accompanied by loss of association to H3K9me2 (Costa-Nunes, Pontes, Preuss, & Pikaard, 2010), suggesting that H3K9me2 is associated with epigenetic silencing of rDNA genes. In order to conclude whether H3K9me2 is associated with rDNA in our cell culture, the use of rDNA FISH in combination with the H3K9me2 localization would allow a more direct comparison.

The H3S10p nuclear localization pattern is similar to the pattern seen for H3K9me2, but the signal in the nucleolus appears to be similar to previously described “nucleolar vacuoles”. These regions of the nucleolus have been proposed as sites of pre-rRNA processing (Beven et al., 1996), but their function remains poorly understood (Raska, Shaw, & Cmarko, 2006). Indeed, the pattern we see is strikingly similar to the distribution of
H3S10p in the nucleoli of tobacco leaf mesophyll cells reported by Granot, et al (2009). That pattern was reported to be similar to the patterns seen for fibrillar centers where the ribosomal RNA is synthesized (Granot et al., 2009). In addition to observing the subnuclear distribution of H3S10p, Granot, et al (2009) used ChIP to confirm that H3S10p is localized on ribosomal DNA (rDNA). Granot, et al (2009) proposed that the nucleolar H3S10p was localized to transcriptionally active regions of the nucleolus, in addition to its well-known role of condensing chromosomes in mitosis.

We find that H3S10p subnucleolar distribution is absent or not perceptible in G1 and S phase nuclei, but is present in nuclei from G2/ early M, which makes it hard to correlate with Granot’s study because it is not expected that the rDNA is only replicated during G2/early M phase. However, we also cannot rule-out the possibility that the H3S10p localization patterns that we see are the initiation of chromosome condensation that peaks in metaphase.

In summary, we find no evidence for large changes in the abundance of different modified histones, as seen by immunoblot analysis, in different cell cycle stages. Similarly, at the level of subnuclear localization, we find no significant changes in the patterns seen for H4K16ac and H3K56ac. However, the pattern for H3S10p is clearly different as it is present in nucleoli during G2/early M phase. Our data are consistent with the idea that H3S10p has multiple roles, which include its well-studied role in mitotic chromosome condensation and potential a role within the nucleolus, where it is found primarily during G2/early M phase.
REFERENCES


**FIGURES AND TABLES**

**Table 1. Antibodies use.** For $^1$ Western blots and $^2$ immunolocalization

<table>
<thead>
<tr>
<th>Mark</th>
<th>Company</th>
<th>Host</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^2$H3K27me3</td>
<td>Upstate 07-449</td>
<td>Rabbit</td>
</tr>
<tr>
<td>$^2$H3K9me2</td>
<td>Upstate 07-212</td>
<td>Rabbit</td>
</tr>
<tr>
<td>$^{1,2}$H3K56ac</td>
<td>Upstate 07-677</td>
<td>Rabbit</td>
</tr>
<tr>
<td>$^{1,2}$H3S10p</td>
<td>Upstate 06-570</td>
<td>Rabbit</td>
</tr>
<tr>
<td>$^{1,2}$H4K16ac</td>
<td>Upstate 07-329</td>
<td>Rabbit</td>
</tr>
<tr>
<td>$^2$H4K5ac</td>
<td>Santa Cruz Sc-34264</td>
<td>Rabbit</td>
</tr>
<tr>
<td>$^{1,2}$H3CT</td>
<td>Santa Cruz Sc-8654</td>
<td>Goat</td>
</tr>
<tr>
<td>$^1$anti goat 800CW</td>
<td>Li-COR 926-32214</td>
<td>Donkey</td>
</tr>
<tr>
<td>$^1$anti rabbit 680</td>
<td>Li-COR 926-32223</td>
<td>Donkey</td>
</tr>
<tr>
<td>$^2$anti goat 633</td>
<td>Invitrogen A21082</td>
<td>Donkey</td>
</tr>
<tr>
<td>$^2$anti rabbit 633</td>
<td>Invitrogen A21206</td>
<td>Donkey</td>
</tr>
</tbody>
</table>
Table 2. Paired T-test analysis of modified histones in G1, S and G2/early M nuclei

<table>
<thead>
<tr>
<th>Mark</th>
<th>Paired data</th>
<th>T-Test</th>
<th>DF</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biorep 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H4K16ac</td>
<td>S-G1</td>
<td>0.5287</td>
<td>3</td>
<td>0.6336</td>
</tr>
<tr>
<td></td>
<td>G2-S</td>
<td>1.7196</td>
<td>3</td>
<td>0.184</td>
</tr>
<tr>
<td></td>
<td>G2-G1</td>
<td>1.9552</td>
<td>3</td>
<td>0.1455</td>
</tr>
<tr>
<td>Biorep 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H4K16ac</td>
<td>S-G1</td>
<td>1.9306</td>
<td>3</td>
<td>0.1491</td>
</tr>
<tr>
<td></td>
<td>G2-S</td>
<td>-0.9732</td>
<td>3</td>
<td>0.4022</td>
</tr>
<tr>
<td></td>
<td>G2-G1</td>
<td>2.849</td>
<td>3</td>
<td>0.0651</td>
</tr>
<tr>
<td>Biorep 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3K56ac</td>
<td>S-G1</td>
<td>-1.88</td>
<td>3</td>
<td>0.1567</td>
</tr>
<tr>
<td></td>
<td>G2-S</td>
<td>1.8926</td>
<td>3</td>
<td>0.1548</td>
</tr>
<tr>
<td></td>
<td>G2-G1</td>
<td>-1.6859</td>
<td>3</td>
<td>0.1904</td>
</tr>
<tr>
<td>Biorep 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3K56ac</td>
<td>S-G1</td>
<td>-2.562</td>
<td>3</td>
<td>0.0830</td>
</tr>
<tr>
<td></td>
<td>G2-S</td>
<td>0.1703</td>
<td>3</td>
<td>0.8756</td>
</tr>
<tr>
<td></td>
<td>G2-G1</td>
<td>-2.4525</td>
<td>3</td>
<td>0.0914</td>
</tr>
<tr>
<td>Biorep 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3S10p</td>
<td>S-G1</td>
<td>2.0517</td>
<td>3</td>
<td>0.1325</td>
</tr>
<tr>
<td></td>
<td>G2-S</td>
<td>-1.7219</td>
<td>3</td>
<td>0.1836</td>
</tr>
<tr>
<td></td>
<td>G2-G1</td>
<td>2.1997</td>
<td>3</td>
<td>0.1152</td>
</tr>
<tr>
<td>Biorep 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3S10p</td>
<td>S-G1</td>
<td>3.8189</td>
<td>3</td>
<td>0.0316</td>
</tr>
<tr>
<td></td>
<td>G2-S</td>
<td>1.7073</td>
<td>3</td>
<td>0.1863</td>
</tr>
<tr>
<td></td>
<td>G2-G1</td>
<td>3.384</td>
<td>3</td>
<td>0.0399</td>
</tr>
</tbody>
</table>

T test = observed t value. DF = degrees of freedom.
**Table 3. Antibodies dilutions.** Antibody dilution for Western blots and for microscopy. AB=antibody.

<table>
<thead>
<tr>
<th>Mark</th>
<th>Western Blot</th>
<th>Microscopy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1&lt;sup&gt;st&lt;/sup&gt; AB</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; AB</td>
</tr>
<tr>
<td>H4K16ac</td>
<td>1:2000</td>
<td>1:3000</td>
</tr>
<tr>
<td>H4K5ac</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>H3K56ac</td>
<td>1:6000</td>
<td>1:6000</td>
</tr>
<tr>
<td>H3K9me2</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>H3S10p</td>
<td>1:200</td>
<td>1:500</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>H3CT</td>
<td>1:200</td>
<td>1:500</td>
</tr>
</tbody>
</table>
Figure 1. Immunoblot for modified histones. Immunoblot comparing the relative abundance in G1, S, and G2/M of histones H3 or H4 carrying the indicated post translational marks.
Figure 2. Confocal microscopy images of sorted Arabidopsis nuclei in G1, S or G2/early M phase. The first column shows nuclei stained with DAPI. Immunochemical signal for the indicated post-transcriptional mark is shown in the second column. The third column shows the signals obtained with the H3CT (sc-8654) antibody to total histone H3. Merged images of the total histone H3 and the PTM signal are shown in the fourth column, and the fifth column shows three-way merged images of DAPI, PTM and total histone H3. Antibodies used are in Table 1.
A) H4K5ac

<table>
<thead>
<tr>
<th></th>
<th>DAPI</th>
<th>Mark</th>
<th>H3 CT</th>
<th>Merged</th>
<th>Merged</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### B) H4K16ac

<table>
<thead>
<tr>
<th>Phase</th>
<th>DAPI</th>
<th>Mark</th>
<th>H3 CT</th>
<th>Merged 1</th>
<th>Merged 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td>S</td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
<tr>
<td>G2</td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
</tr>
<tr>
<td></td>
<td>DAPI</td>
<td>Mark</td>
<td>H3 CT</td>
<td>Merged</td>
<td>Merged</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>-------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td><strong>G1</strong></td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>S</strong></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>G2</strong></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
</tr>
</tbody>
</table>
D) **H3K9me2**

<table>
<thead>
<tr>
<th></th>
<th>DAPI</th>
<th>Mark</th>
<th>H3 CT</th>
<th>Merged</th>
<th>Merged</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>G1</strong></td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
<td><img src="image5" alt="Image" /></td>
</tr>
<tr>
<td><strong>S</strong></td>
<td><img src="image6" alt="Image" /></td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td><img src="image9" alt="Image" /></td>
<td><img src="image10" alt="Image" /></td>
</tr>
<tr>
<td><strong>G2</strong></td>
<td><img src="image11" alt="Image" /></td>
<td><img src="image12" alt="Image" /></td>
<td><img src="image13" alt="Image" /></td>
<td><img src="image14" alt="Image" /></td>
<td><img src="image15" alt="Image" /></td>
</tr>
</tbody>
</table>
### H3K27me3

<table>
<thead>
<tr>
<th></th>
<th>DAPI</th>
<th>Mark</th>
<th>H3 CT</th>
<th>Merged</th>
<th>Merged</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td><img src="G1_DAPI.png" alt="Image" /></td>
<td><img src="G1_Mark.png" alt="Image" /></td>
<td><img src="G1_H3_CT.png" alt="Image" /></td>
<td><img src="G1_Merged.png" alt="Image" /></td>
<td><img src="G1_Merged.png" alt="Image" /></td>
</tr>
<tr>
<td>S</td>
<td><img src="S_DAPI.png" alt="Image" /></td>
<td><img src="S_Mark.png" alt="Image" /></td>
<td><img src="S_H3_CT.png" alt="Image" /></td>
<td><img src="S_Merged.png" alt="Image" /></td>
<td><img src="S_Merged.png" alt="Image" /></td>
</tr>
<tr>
<td>G2</td>
<td><img src="G2_DAPI.png" alt="Image" /></td>
<td><img src="G2_Mark.png" alt="Image" /></td>
<td><img src="G2_H3_CT.png" alt="Image" /></td>
<td><img src="G2_Merged.png" alt="Image" /></td>
<td><img src="G2_Merged.png" alt="Image" /></td>
</tr>
</tbody>
</table>
F) H3S10P

<table>
<thead>
<tr>
<th></th>
<th>DAPI</th>
<th>Mark</th>
<th>H3 CT</th>
<th>Merged</th>
<th>Merged</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td><img src="G1_DAIF.png" alt="Image" /></td>
<td><img src="G1_MARK.png" alt="Image" /></td>
<td><img src="G1_H3_CT.png" alt="Image" /></td>
<td><img src="G1_Merged.png" alt="Image" /></td>
<td><img src="G1_Merged.png" alt="Image" /></td>
</tr>
<tr>
<td>S</td>
<td><img src="S_DAIF.png" alt="Image" /></td>
<td><img src="S_MARK.png" alt="Image" /></td>
<td><img src="S_H3_CT.png" alt="Image" /></td>
<td><img src="S_Merged.png" alt="Image" /></td>
<td><img src="S_Merged.png" alt="Image" /></td>
</tr>
<tr>
<td>G2</td>
<td><img src="G2_DAIF.png" alt="Image" /></td>
<td><img src="G2_MARK.png" alt="Image" /></td>
<td><img src="G2_H3_CT.png" alt="Image" /></td>
<td><img src="G2_Merged.png" alt="Image" /></td>
<td><img src="G2_Merged.png" alt="Image" /></td>
</tr>
</tbody>
</table>
CHAPTER IV

Prospectus
Accurate replication of the DNA during the cell cycle is critical for cell survival. The first model for DNA replication proposed that an initiator protein binds to a specific DNA sequence (origin) to start DNA replication (Jacob & Brenner, 1963). This model held true for bacteria, virus and yeast (S. cerevisiae) where it has been demonstrated that origins are sequence specific. However, DNA replication in higher eukaryotes is more complex and a specific sequence for origins of DNA replication remains elusive. However, while a specific sequence has yet to be reported, some origins from higher eukaryotes have been well characterized. Unfortunately the origins of DNA replication in plants remains poorly understood, as only one origin from ribosomal DNA has been well characterized (Van't Hof et al., 1987).

The goal of my research has been to locate the origins of DNA replication on the AtCh4 DLA of Arabidopsis thaliana (Col-0). I attempted several approaches that have been used by others for locating origins of replication from other eukaryotes, but soon realized that the use of such approaches has serious limitations for plant studies. In particular, a major limitation that I faced was having a well-defined origin of replication from A. thaliana to use as a positive control for my experiments. Here I propose future directions for mapping origins in plants along with a proposal for studies to explore the changes that take place in the post-translational (PTM) histone modifications as the cell progresses through the cell cycle.
The majority of my project required the use of the small nascent strands (SNS) that are produced during DNA replication. During my project new technical advances were developed for improving nascent strand purification, which should be incorporated into future attempts to map origins of replication in plants.

A major example would be the use of an alternative approach where EdU (5-ethynyl-2’-deoxyuridine) (Amiard et al., 2010) replaces BrdU and subsequent sorting of the nuclei using DNA content (DAPI) as described in Chapter 2. The use of EdU has several major advantages over BrdU, one being that the fluorescent probe used to detect EdU is a small permeable molecule that allows to it easily target the EdU in the cell (Cavanagh, Walker, Norazit, & Meedeniya, 2011). To improve my experiments, I would sort nuclei in S phase to enrich for active origins of replication, the sorting would be based on DNA content, avoiding the use of BrdU, which would greatly reduce possible background that might occur from BrdU-caused DNA breakage and repair, prior to, and after harvesting.

Unfortunately, a major limitation of sorted material coupled with SNS-based origin of replication mapping is the initial amount of material needed. For example even when asynchronous cells are used to isolate nascent strands the Exo-SNS procedure yields a vanishingly small amount of nascent strands as the starting material from mapping.

In order to use sorted nuclei for such experiments an adequate number of nuclei would need to be collected by FACS. To carryout the Exo SNS, the DNA is first isolated from the nuclei, followed by size selection for the exo-nuclease digestion. Cayrou, et al (2011) describe the use of a double incubation with a “customized lambda exonuclease” from Fermentas, Inc. as the key step for obtaining a more highly purified SNS sample when
compared to the “SNS” when other sources of lambda exonuclease are used (Cayrou, Coulombe et al. 2011). We should point out that Cadoret, et al (2008) have estimated that 30,000 origins are activated inside a given human cell which yields approximately 10 ng of SNS from $10^8$ exponentially growing cells for origin mapping when the Exo SNS procedure (with Fermentas enzyme) is used (Cadoret et al., 2008). If the use of sorted material is not possible due to the limiting amount of SNS that can be isolated, the FACS step might be eliminated to determine whether the Fermentas lambda exonuclease treatment results in a level of purity to make the sorting unnecessary (Cadoret et al., 2008; Cayrou et al., 2011).

Another approach would be to take advantage of the stability of EdU in the presence of UV, along with using the DAPI-stained DNA for sorting the nuclei by DNA content. This approach is a simple modification of our original approach for the BrdU SNS. To isolate EdU-labeled fragments, DNA containing EdU is first conjugated to biotin using the “click it” reaction and precipitated using streptavidin-coated beads to isolate the nascent EdU-labeled DNA (Sirbu et al., 2011).

A modification of the EdU-biotin conjugation approach would take advantage of the biotin-streptavidin binding affinity (Suter, Cazin, Butler, & Mock, 1988). For this approach, the nascent strands would be size-selected, and affinity purified using the EdU-Biotin conjugate. Lambda exonuclease would then be added with the EdU-Biotin conjugated DNA remaining bound to the streptavidin resin using the ChIP-exo procedure described by Rhee and Pugh (Rhee & Pugh, 2011). The lambda exonuclease digests the unprotected 5’ DNA that are not protected by the RNA primer. The short sequences, labeled with EdU and
containing the RNA primer will provide a highly purified source of nascent strands for subsequent sequencing, or for hybridization to a microarray for mapping.

Validation of any result acquired by high throughput methods is crucial. Therefore, it would be best to use 2-D gel analysis, which is considered to be the “gold standard” for the analysis of origins of replication (Aparicio, personal communication). The combination of a highly efficient method to isolate highly purified SNS, and a confirmation by 2-D gels, may allow researchers to characterize well-defined origins of DNA replication from the Arabidopsis genome. Such data can then be used in a variety of ways such as mapping the origins of replication in tissues or even whole organisms. The origin mapping data can be compared with existing annotations and epigenetic maps such as histone modifications, DNA methylation, gene expression, and etcetera.

By comparing the genome locations where genomic and epigenomic features converge; we will gain a deeper understanding of how origins of replication are selected and how DNA replication is regulated. Understanding the origins of DNA replication in plants can eventually help in designing artificial chromosomes, which may improve plant transformation by allowing the simultaneous introduction of several genes while avoiding expression variation due to differences in the chromatin structure at site of transgene integration. When combined with information about transcription and chromatin structure, this information will help clarify the interplay between DNA replication and plant development.
The Dynamic Epigenome

My study has revealed that any changes in the levels in the bulk pool of histone PTMs that may occur during the transition from G1, S, and G2/early M cannot be measured by Western blot analysis. However, the preliminary microscopy results indicate that some of the histone PTMs that we examined show changes in subnuclear distribution. We found that the location of some histone PTMs, such as H3S10p, also have phase-specific changes that occur within the nucleolus. Future experiments designed to understand the role of nucleolar H3S10p during the transition from late G2 to M are a promising area of research, since there is little known about that epigenomic changes that take place in the nucleolus during DNA replication and cell division. Such studies of the nucleolus have been difficult because of the repetitive DNA, which makes mapping difficult.

Recently Sirbu, et al (2011) developed a procedure they called “iPond” (isolation of proteins on nascent DNA), in which nuclei are pulse-labeled with EdU, which is then conjugated with biotin, allowing nascent DNA and its associated proteins to be captured as an EdU-biotin complex on streptavidin beads. Importantly the procedure allowed the chromatin-associated proteins to remain intact (Sirbu et al., 2011). The authors then either varied the EdU labeling times, or followed the EdU labeling with a chase of thymidine for varying times, in order to determine when chromatin structure was re-established following DNA replication. The use of the iPond and analogues such as EdU will allow researchers to address questions that, until now, have been intractable.

In our study we chose to study the histone PTMs that have been shown to play a role in replication in other higher eukaryotes. While this approach is logical, the use of EdU
may help us to make such decisions based upon the changes of the histone PTM composition at the Arabidopsis replication fork. To do such an experiment the genomic DNA would be pulse labeled with EdU, nuclei sorted into G1, S and G2/early M, the conjugated EdU-biotin chromatin captured on the streptavidin beads, for analysis of the associated proteins by proteomic analysis. Recently Kliszczak, et al (2011) demonstrated the feasibility of using this approach for proteomic analysis to provide direct information about the histone PTMs along with the additional replication-associated proteins (Kliszczak, Rainey, Harhen, Boisvert, & Santocanale, 2011). Such a study would be extremely valuable for understanding plant DNA replication because until now project designs have been based upon the replication proteins associated with other higher eukaryotic organisms, which may have some differences in plants (Shultz et al., 2007). The new data could then serve as a foundation for choosing the important histone PTMs, or other proteins, that are associated with active replication for a more detailed analysis.

The use of EdU for labeling nuclei also has great advantages for high-resolution microscopy. Nuclei containing EdU-labeled DNA could be probed with the fluorescent antibodies of the key histone PTMs of interest or other replication-associated proteins (Cseresnyes, Schwarz, & Green, 2009), possibly discovered from the proteomics analysis, to determine whether co-localization occurs with newly synthesized DNA in specific regions within the nucleus.
REFERENCES


