Genome-wide analysis of nucleosome occupancy surrounding *Saccharomyces cerevisiae* origins of replication

by

Nicolas Matthew Berbenetz

A thesis submitted in conformity with the requirements
for the degree of Master of Science

Molecular Genetics
University of Toronto

Genome-wide analysis of nucleosome occupancy surrounding *Saccharomyces cerevisiae* origins of replication

Nicolas Matthew Berbenetz

Master of Science

Molecular Genetics
University of Toronto

2010

Abstract

The *Saccharomyces cerevisiae* origin recognition complex (ORC) binds to replication origins at the ARS consensus sequence (ACS), serving as a scaffold for the assembly of replication complexes needed for the initiation of DNA synthesis. I generated a genome-wide map of nucleosome positions surrounding replication origins because the precise locations of nucleosomes may influence replication. My map revealed a nucleosome-free region surrounding the ACS that is bordered by two well-positioned nucleosomes. I was able to explain differences in origin properties by clustering nucleosome profiles. I found an association between the replication time and nucleosome profile for a given origin cluster. An ORC depletion mutant nucleosome map indicated a shift in nucleosomes towards the ACS. I present the first genome-wide view of origin nucleosome architecture, indicate a relationship between chromatin structure and replication timing, and suggest a model whereby the interplay between DNA sequence and ORC binding defines the nucleosome occupancy pattern.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACS	ARS consensus sequence
ARS	Autonomously Replicating Sequence
bp	base pairs
CDK	cyclin-dependent kinase
ChIP	chromatin immunopreciptation
DNA	Deoxyribonucleic acid
DNase	deoxyribonuclease
HU	hydroxyurea
MCM	mini-chromosome maintenance
NDR	Nucleosome depleted region
NPS	Nucleosome positioning sequence
ORC	Origin recognition complex
ORF	Open reading frame
pre-RC	Pre-replicative complex
PWM	Position weight matrix
TF	Transcription Factor
TSS	Transcription start site

Chapter 1
Introduction

1.1 Genome-wide analysis of nucleosome locations

1.1.1 An introduction to the nucleosome

DNA metabolic processes occur in the context of chromatin. The basic level of chromatin is a repeating structure with DNA wrapped 1.7 turns around histone core particles or nucleosomes. Since the proposal of the "beads on a string" model of nucleosomes in the 1970s (Kornberg, 1974) there has been steady progress in our understanding of how nucleosome positions affect fundamental biological processes in eukaryotes. In the past couple of years advances in yeast genomics have led to a better understanding of nucleosome positioning in higher organisms.

In eukaryotes, genomic DNA is not freely accessible but rather is bound to histone proteins and packaged. The nucleosome hypothesis described the basic repeating unit of chromatin as a segment of DNA wrapped around histone proteins (Kornberg, 1974). This hypothesis explained the existing x-ray diffraction patterns of chromatin, the stoichiometry of histones and DNA, as well as the laddering of chromatin digested with micrococcal nuclease (Kornberg, 1974). The nucleosome hypothesis was confirmed through the determination of a high-resolution X-ray crystal structure of the nucleosome core particle (Luger et al., 1997), which consists of 147-bp of DNA wrapped around a histone octamer composed of two molecules each of the histone proteins: H2A, H2B, H3 and H4 (Luger et al., 1997). The histone octamer surface is positively charged and superhelical, allowing DNA to be wrapped in a superhelix of approximately 1.65 turns with 10.2-bp per turn (Luger et al., 1997).

As soon as the nucleosome model was proposed, it raised the question of whether specific DNA sequences preferentially bound nucleosomes (Kornberg, 1974). Early ideas suggested that

nucleosome positioning can be a consequence of statistical positioning in which a strong DNA-protein interaction acts as a boundary and leads to the formation of an array of positioned nucleosomes extending away from the boundary (Kornberg, 1981). Alternatively, nucleosome positioning could be sequence encoded; sequences with high histone octamer affinity would be expected to be found within nucleosomes preferentially (Simpson, 1986). This model predicts that the DNA sequence itself encodes all nucleosome locations (Ioshikhes et al., 2006; Segal et al., 2006). Recent models of nucleosome occupancy in eukaryotes incorporate both concepts (Jiang and Pugh, 2009).

Nucleosome positioning influences all biochemical processes in which DNA is involved, e.g., recombination and DNA damage repair, replication, and transcription (Luger et al., 1997). This is a consequence of nucleosomes influencing the accessibility of trans acting factors to DNA. DNA within the linker regions that lie between nucleosomes is fully accessible while nucleosomal DNA is only partially accessible (Simpson, 1986). Nucleosomes are not limited to influencing DNA-protein interactions. Their histone tails, which protrude from the core particle, are subject to multiple post-translational modifications. These tails can recruit proteins leading to chromatin remodelling which can either activate or repress DNA metabolic processes (Segal et al., 2006).

1.1.2 Overview of methods to determine nucleosome positions

The recent surge in chromatin-focussed research is a consequence of studies indicating the influence of histone mutations on chromatin structure and the importance of chromatin remodelling proteins in gene expression studies, combined with new genomic technologies (Rando, 2007; Simpson, 1999). Before genome-wide information on nucleosome positions in yeast was available, knowledge was limited to single gene studies performed *in vitro* and *in vivo*.

The main tool to detect *in vivo* positioned nucleosomes has not changed: it involves using a nuclease that preferentially digests chromatin at linker regions. The main difference between the pre-genomic and genomic experiments involves the process to identify nucleosomes. Early studies used restriction enzyme digests of nuclease-treated chromatin followed by Southern blotting in order to identify nucleosomes (Simpson, 1986). Sites cut in chromatin and genomic DNA are linker regions, if the distance between two linkers was larger than the length of a nucleosome repeat (147-bp) the DNA segment was considered nucleosomal (Simpson, 1986). Current studies rely on high-throughput DNA sequencing or microarray hybridization in order to detect nucleosome locations (Jiang and Pugh, 2009). Another difference between pre-genomic and genomic studies involves the use of formaldehyde to fix chromatin so that interactions between histones and DNA are maintained (Simpson, 1999).

Pre-genomic studies of nucleosome positioning revealed that nucleosome locations can be random or precisely localized (Kornberg and Lorch, 1992). Positioned nucleosomes can interfere with DNA metabolic processes, for example, the repression of *S. cerevisiae MATa*-specific genes such as *STE6* by *MATα2* (expressed by *MATα* cells) is a result of nucleosomes being positioned over the promoter and transcription start site in *MATα* cells but not in *MATa* cells (Shimizu et al., 1991). The positioning of these nucleosomes was established by performing primer-extension on micrococcal nuclease treated chromatin from *MATα* and *MATa* cells (Shimizu et al., 1991).

The earliest genome-wide study of nucleosome positions was performed using Simian Virus 40 (SV40) (Ambrose et al., 1990). By cloning micrococcal nuclease digested SV40 fragments into a vector it was possible to identify the precise locations of nucleosomes within the SV40 genome. By counting the number of sequences for each position in the SV40 genome it was possible to obtain nucleosome density information which revealed alternating regions of high and low

nucleosome occupancy (Ambrose et al., 1990). Nucleosome locations were identified and classified into three groups: strong, weak and randomly positioned, based on the proximity and number of nucleosome midpoint calls (Ambrose et al., 1990). The strongest positioned nucleosome was found within 8-bp of the main SV40 late gene transcription start site. Other strongly positioned nucleosomes were found in different late genes, while, early genes contained randomly positioned nucleosomes (Ambrose et al., 1990). Presumably, the lack of positioned nucleosomes allows the expression of early genes without nucleosome interference. The method introduced by this paper to identify nucleosome locations is currently used to identify nucleosomes in other organisms. The main improvement involves the direct, high-throughout sequencing of micrococcal nuclease digested DNA, i.e., without DNA cloning.

1.1.3 DNA-encoded nucleosome locations

A significant finding during the pre-genomic era was that certain DNA sequences were preferentially nucleosome bound. For example, histone octamers from different species (e.g. chicken, yeast, human, etc.) bind *in vitro* to specific sequences within the 5S rRNA gene generating a positioned nucleosome (Hayes and Wolffe, 1992). The precise nucleosome positioning signal of 5S rRNA was within the central ~60-bp of DNA bound by the histone octamer (FitzGerald and Simpson, 1985). This positioned nucleosome covers the 5S rRNA transcription start site and prevents transcription by restricting access to the TFIIIA transcription factor binding site (Hayes and Wolffe, 1992). Transcription of 5S rRNA occurs when the TFIIIA binding site is exposed following the acetylation of histone (H3/H4) tails contained within the nucleosome positioned over the 5S rRNA transcription start site (Lee et al., 1993). In general, it is possible to identify DNA sequences preferentially incorporated into nucleosomes by observing a 10-bp periodicity in the laddering of fragments produced following DNase I digestion of radiolabelled, well-positioned, nucleosomal DNA (Simpson, 1986).

Several *in vitro* studies demonstrated that any DNA sequence could be nucleosomal but certain sequences, dubbed nucleosome-positioning sequences, have a greater tendency to be nucleosomal (Thastrom et al., 1999; Widom, 2001). This result is explained by different DNA sequences having different energy requirements to form a nucleosome; this energy is needed to bend, twist and melt DNA (Widom, 2001). A large portion of the chemical energy gained from histone-DNA interactions is used to bend DNA within the nucleosome (Widom, 2001). In solution 150-bp DNA segments tend to be straight while longer lengths of DNA are bent (Widom, 2001). Furthermore, DNA within the nucleosome is sharply bent every 5-bp within the 10-bp helical repeat of DNA within a nucleosome: first, when the major groove contacts the histone octamer and second, when the minor groove contacts the histone octamer (Luger et al., 1997). Based on *in vitro* studies GC-rich sequences are expected when the minor groove faces the histone octamer, and AT-rich sequences are expected when the major groove faces the histone octamer (Thastrom et al., 1999). Thus, DNA sequences containing AT- and GC-rich bases at sites which are sharply bent within the nucleosome have the highest nucleosome affinity and form the most stable nucleosomes (Widom, 2001).

Nucleosome positioning refers to the average location of nucleosomes within a population of cells. All possible positions along a DNA sequence can be nucleosome occupied, but in an average view of nucleosome positioning only the most preferred sequences are occupied (Thastrom et al., 1999). Nucleosome positioning is characterized by translational positioning, selecting a particular 147-bp tract of DNA as opposed to other tracts obtained by sliding (short-range nucleosome movements) forwards or backwards along the DNA, and rotational positioning, a set of sequences obtained by sliding forwards or backwards by 10-bp (the helical repeat length of DNA within a nucleosome) in order to maintain the orientation of specific DNA bases with the histone octamer (Thastrom et al., 1999). DNA within the nucleosome interacts

(through hydrogen bonds and salt bridges) with the histone octamer at 14 sites, generating a stable structure (Luger et al., 1997). Rotational positioning changes (~10-bp movements) of the nucleosome can occur passively by disrupting one histone-DNA interaction at the end of the nucleosome followed by the formation of a new interaction with a different base and the formation of a temporary bulge of DNA (Becker, 2002). This bulge (bent DNA) diffuses to the other end of the nucleosome, disrupting one histone-DNA interaction at a time leading to the translocation of the histone octamer relative to the underlying DNA (Becker, 2002). Moving nucleosomes over larger distances (up to 100-bp) requires the use of ATP-dependent chromatin remodellers (Chou, 2007). ATP-dependent chromatin remodellers can catalyze the sliding of nucleosomes or the complete removal of a histone octamer from a segment of DNA (Becker, 2002).

A nucleosome positioning code was recently proposed (Ioshikhes et al., 2006; Segal et al., 2006). Segal et al. sequenced ~200 yeast nucleosomal DNA sequences and determined nucleosome sequence preferences using DNA dinucleotide distributions, which capture differences in DNA bending. They found that AA/TT/TA dinucleotides are preferred at the nucleosomal DNA minor groove when DNA is in contact with histones while GC is preferred at the minor grove when nucleosomal DNA is at its furthest distance to histones ~5-bp away (Segal et al., 2006). Using sequenced nucleosomal DNA, Segal et al. were able to predict the locations of nucleosomes genome-wide. Using a set of ~100 nucleosomes identified in previous studies, their model was able to predict ~50% of nucleosomes within 35-bp of their reported positions (Segal et al., 2006). Nucleosomes tend to occupy transcription factor binding sites, leaving only a small proportion available for transcription factors (Segal et al., 2006). The ability of certain nucleosomes to be remodelled may be sequence encoded by specifying low affinity nucleosomes over a particular region (Segal et al., 2006). This result contradicts the expectation that nucleosome sequence

preferences are not relevant due to the presence of ATP-dependent chromatin remodellers (Ercan and Lieb, 2006), which can move nucleosomes to non-preferred sequences (Segal et al., 2006).

Ioshikhes et al. (2006) developed a complementary model of sequence-encoded nucleosome positioning. They examined a set of co-regulated genes from a histone H4 deacetylase mutant and compared nucleosome positioning sequence correlation to a collection of ~200 well-positioned nucleosomes. TATA-less (80% of genes) and TATA-containing (20% of genes) promoters had distinct nucleosome positioning sequence arrangements (Ioshikhes et al., 2006). Correlation peaks corresponded to predicted nucleosome locations while troughs corresponded to a nucleosome free region or linker (Ioshikhes et al., 2006). Ioshikhes et al. were able to generate a model based on orthologous nucleosomal DNA sequences from related *Saccharomyces* species and were able to predict the location of known nucleosome positions experimentally derived for chromosome 3 (Yuan et al., 2005). Clustering individual genes based on their nucleosome positioning sequence correlation revealed an NPS-NDR-NPS pattern at promoters (Ioshikhes et al., 2006). The studies by Ioshikhes et al. and Segal et al. indicate that DNA sequence is one determinant of nucleosome positioning in genomes. The diffuse nucleosome positioning signal identified by Ioshikhes et al. and Segal et al. provides an explanation for 15-20% of nucleosome positions in the genome (Shivaswamy et al., 2008; Zhang et al., 2009).

The existence of positioned nucleosomes poses an interesting paradox; nucleosome-bound DNA is thought to be inaccessible to DNA metabolic processes including recombination, repair, replication, and transcription, yet these processes occur despite the presence of positioned nucleosomes (Anderson and Widom, 2000; Pazin et al., 1997). This paradox can be partially resolved without invoking ATP-dependent chromatin remodellers in the “site exposure model” which posits that the DNA within a nucleosome is in equilibrium with translationally moved

(sliding nucleosomes) or uncoiled (where DNA is unwrapped in 10-bp increments while the rest of the DNA sequence remains bound to the histone octamer) nucleosomes (Anderson and Widom, 2000). Thus, any DNA sequence within a positioned nucleosome is potentially accessible depending upon the affinity between DNA and histone octamer within a nucleosome (Anderson and Widom, 2000). However, to enhance the rate of site-exposure, chromatin-remodellers are required. Together, transient site-exposure and chromatin remodellers resolve the paradox of why positioned nucleosomes do not render DNA inaccessible. Transient site-exposure and the statistical positioning of nucleosome model could explain why the locations of positioned nucleosomes change when a gene is activated or repressed (Pazin et al., 1997). During the transient exposure of a transcription factor binding site, a transcription factor can create a barrier which positions adjacent nucleosomes. Once the transcription factor is no longer bound, nucleosomes reposition themselves to their most thermodynamically preferred arrangement (Pazin et al., 1997).

1.1.4 Genome-wide nucleosome maps

Accessibility to DNA regulatory-sites such as transcription factor binding sites is dependent upon the location of nucleosomes. An early indication of the importance of nucleosome positioning came from a study using low resolution microarrays (constructed with long PCR amplicons) which found promoters to be nucleosome-depleted relative to ORFs (Lee et al., 2004). A study by Yuan et al. provided the first high-resolution view of nucleosome positions. Yuan et al. developed a microarray approach to identify nucleosomes based on the susceptibility of linker DNA to micrococcal nuclease digestion. Nucleosome positions were identified by isolating nucleosomal DNA and genomic DNA followed by competitive hybridization to a tiling array comprised of 60 nucleotide probes that overlapped and covered chromosome 3 (Yuan et al., 2005). Yuan et al. identified nucleosome positions as peaks in log2 transformed hybridization

signal (nucleosomal vs. genomic DNA) with troughs corresponding to linkers. Using a hidden Markov model, they were able to classify ~69% of chromosome 3 DNA as occupied with well positioned nucleosomes (which cover ~147bp) while the remaining sequence was covered by fuzzy nucleosomes (covering more than ~147bp) or completely unoccupied (i.e., a linker region) (Yuan et al., 2005). Yuan et al. confirmed that promoters tend to be nucleosome depleted (Lee et al., 2004) and determined a pattern of nucleosome occupancy at coding genes: a nucleosome-free region of ~150-bp encompassing the transcriptional start site bordered on either side (intergenic and in the direction of the ORF) by well-positioned nucleosomes (the -1 and +1 nucleosomes). The significance of positioned nucleosomes was revealed by the determination that the majority (87%) of motifs associated with transcription factors were in nucleosome-free regions or linkers (Yuan et al., 2005). Finally, the importance of nucleosome positioning sequences was revealed by the observation that nucleosome-depleted regions (NDRs) which contain rigid poly(dA:dT) tracts have poor nucleosome affinity (Yuan et al., 2005).

The nucleosome positions identified by Yuan et al. were used to predict genome-wide nucleosome locations computationally (Peckham et al., 2007). In contrast to previous models (Ioshikhes et al., 2006; Segal et al., 2006) the Peckham et al. model predicts that not all nucleosomes are DNA encoded. The strongest known, eukaryotic nucleosome positioning sequences (including the well-studied 5S rRNA promoter) are significantly weaker than synthetic sequences, indicating eukaryotic genomes do not take complete advantage of nucleosome positioning sequences (Thastrom et al., 1999). The GC/AT-richness of a given sequence strongly influences its nucleosome positioning potential (Peckham et al., 2007). The Peckham et al. model predicted ~17% more nucleosomes than expected by chance demonstrating that DNA sequence has a subtle influence on the locations of most nucleosomes. Nucleosome

exclusion signals within promoters have a stronger influence on nucleosome positioning than nucleosome positioning motifs within open reading frames (Peckham et al., 2007).

The first genome-wide map of nucleosome locations focussed on identifying the histone variant H2A.Z using high-throughput sequencing (Albert et al., 2007). The high-resolution nucleosome map indicated that transcription factor binding sites occur upstream of the +1 nucleosome (first nucleosome to the right of the transcription start site). The +1 nucleosome border contains the transcription start site within its first helical turn (10-bp) of DNA. Furthermore, conserved transcription factor binding sites reside near nucleosome borders suggesting that transcription factors could translationally displace nucleosomes. Using the locations of H2A.Z nucleosomes, AA/TT and GC dinucleotide periodicities correspond with the thermodynamically preferred arrangement of AA/TT and GC dinucleotides (Albert et al., 2007). Poorly positioned (fuzzy) nucleosomes were defined using the standard deviation of sequencing read coordinates for a particular nucleosome (Albert et al., 2007). Fuzzy nucleosomes were found to contain TATA-boxes and were regulated by chromatin remodellers. Different chromosomal elements such as telomeres, centromeres, origins, and ORFs were found to have distinct nucleosome architectures (Albert et al., 2007). Telomeres contain fixed H2A.Z nucleosomes ~200-bp apart while centromeres lacked any H2A.Z nucleosomes. Origins of replication lack H2A.Z but flanking DNA sequences contain H2A.Z nucleosomes. TATA-less promoters contain H2A.Z nucleosomes flanking the promoter nucleosome-free region while TATA-containing promoters contain fuzzy H2A.Z nucleosomes. The distinct nucleosome architectures of different chromosomal elements could correlate with their function.

The first complete genome-wide nucleosome map was obtained using a tiling microarray with 4-bp resolution (Lee et al., 2007). Using a modification of the Yuan et al. hidden Markov model,

Lee et al. determined that 81% of the yeast genome is covered by nucleosomes: ~40,000 well-positioned and ~30,000 fuzzy nucleosomes. Nucleosome occupancy correlated with transcript abundance and functionally related genes could be grouped together based on their nucleosome occupancy patterns. Transcription factor binding sites were enriched within the promoter nucleosome depleted region. Lee et al. developed a model which explained nucleosome occupancy patterns better than an earlier model (Segal et al., 2006) by incorporating transcription factor binding sites, DNA dinucleotide properties and other factors influencing nucleosome positioning (Lee et al., 2007). Comparing predicted nucleosome locations with experimentally observed nucleosome occupancy the Lee et al. model had a correlation coefficient of 0.44 while the Segal et al. model had a correlation coefficient of 0.09.

A similar genome-wide map was obtained using high-throughput sequencing of immunopurified histones H3 and H4 (Mavrich et al., 2008). In this study, DNA sequence was sufficient to explain the nucleosome-depleted region and its adjacent -1 (intergenic) and +1 (ORF) nucleosomes (Mavrich et al., 2008). Sequence elements influencing the promoter-proximal nucleosomes include nucleosome positioning sequences AA/TT (minor groove) and GC (major groove), nucleosome excluding sequences (rigid poly (dA:dT) tracts), and DNA regulatory sites (transcription factor binding sites) (Mavrich et al., 2008). Distal to the NDR the possible locations that nucleosomes can occupy are limited, leading to increased fuzziness in their positions (Mavrich et al., 2008). Nucleosome fuzziness is based on all sequences found to contribute to a particular nucleosome location. Well-positioned nucleosomes have little translational movement in contrast to poorly-positioned nucleosomes. Both Mavrich et al. and a study by Whitehouse et al. determined the importance of the 3' NDR in transcription termination, in inhibition of anti-sense transcription, and possibly a role in looping the

transcriptional machinery back to the promoter via binding sites for TFIIB (Mavrich et al., 2008; Whitehouse et al., 2007).

In general, the different genome-wide nucleosome maps obtained from wild-type yeast indicated that the organization of nucleosomes fits the model for statistical positioning of nucleosomes (Jiang and Pugh, 2009). Statistical positioning of nucleosomes is a consequence of nucleosomes being arranged in an array of adjacent nucleosomes. By positioning the first nucleosome in an array of nucleosomes the positions of subsequent nucleosomes are affected because of limited lateral mobility of nucleosomes (Kornberg and Stryer, 1988). As the distance from the positioned nucleosome increases nucleosomes are less restricted by adjacent nucleosomes and their positions are increasingly delocalized (**Figure 1**). Furthermore, coding genes have a distinct nucleosome occupancy pattern in which there is a nucleosome-free promoter bracketed by two-well positioned nucleosomes. The intergenic -1 nucleosome and the array of intergenic nucleosomes have poor phasing compared to the transcription start site containing +1 nucleosome. Nucleosomes within the ORF have progressively lower phasing away from the +1 nucleosome. The decrease in phasing fits the statistical positioning of nucleosomes model.

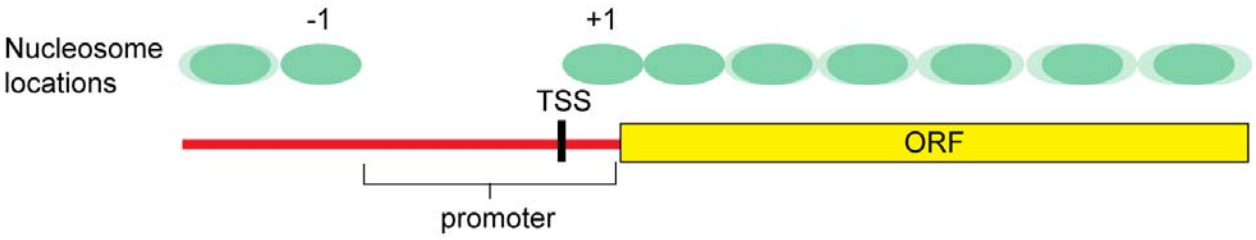

Figure 1: The statistical positioning of coding gene nucleosomes.
The +1 and -1 nucleosomes flank a coding gene promoter. The +1 nucleosome contains the transcription start site (TSS). Further away from the nucleosome-free promoter nucleosomes are progressively more delocalized, indicated by increased delocalization of nucleosome positions. Adapted from Mavrich et al. (2008).

1.1.5 Nucleosome positions are dynamic

Nucleosome positioning has long been suspected to have a role in gene expression. Genome-wide studies on wild-type (S288C) yeast attempted to address this question by inferring positional dynamics by clustering genes, observing distinct nucleosome occupancy patterns, and correlating these patterns with biological function. For example, highly expressed ribosomal protein genes tend to have reduced nucleosome phasing (Mavrich et al., 2008). A direct demonstration of the influence of nucleosome positioning dynamics on gene expression required the use of genetic or physiological perturbation. That is, distinct conditions which influence the expression of specific genes should cause changes in the nucleosome occupancy at these genes.

A study which used genetic perturbation of a chromatin remodelling protein (Isw2) found a significant influence on nucleosome positioning at a subset of genes (Whitehouse et al., 2007). Whitehouse et al. determined that Isw2 repositions nucleosomes into locations with less-favourable nucleosome occupancy preventing the expression of meiosis-specific genes. The degree of repositioning was determined by selecting 400 Isw2-enriched genes. By overlaying the nucleosome maps of wild-type and *isw2* mutants, nucleosomes were found to be repositioned by 15 to 70-bp in the direction of the ORF in the mutant. These nucleosome positions are more favourable leading to the exposure of transcription initiation sites in an *isw2* mutant (Whitehouse et al., 2007). Genes subject to Isw2 remodelling had a +1 nucleosome covering the transcriptional start site preventing transcription (Whitehouse et al., 2007). This study demonstrates that chromatin remodelling influences nucleosome positioning dynamics genome-wide.

A study (Shivaswamy et al., 2008) which used the physiological perturbation of heat shock (which causes an extensive change in gene expression) indicated that not all nucleosome

positioning changes are associated with changes in transcription. Following heat shock, a small group of nucleosomes were displaced by 100-bp or more; these changes in nucleosome occupancy were not limited to genes with significant transcriptional repression or activation (Shivaswamy et al., 2008). Heat shock activated genes tended to have nucleosomes displaced in the direction of the ORF, displacing a nucleosome covering their promoter, permitting the recruitment of transcription factors (Shivaswamy et al., 2008). In contrast, heat shock repressed genes tended to have nucleosomes repositioned in the direction of the promoter resulting in a nucleosome positioned over their promoter region (-200 to +50-bp) preventing transcription (Shivaswamy et al., 2008). This study demonstrates that chromatin remodelling changes associated with gene activation are associated with promoters becoming nucleosome-free while changes associated with gene repression are associated with the appearance of a nucleosome within the promoter.

Yeast may encode the locations of nucleosome and nucleosome-depleted regions within their DNA sequence. Open chromatin architecture, a nucleosome-free promoter, is usually found at essential genes and genes that require consistent expression while closed chromatin architecture, a nucleosome-covered promoter, is found at nonessential genes or condition-dependent genes (Field et al., 2008). Closed chromatin architecture results in promoters which would be expected to have competition between transcription factors and nucleosomes for access to DNA.

1.1.6 *In vitro* nucleosome occupancy maps

Recent nucleosome occupancy investigations have re-examined the strength of the nucleosome positioning code. Field et al. updated the DNA-encoded nucleosome positioning model using full-length mononucleosome sequencing using 454 Life Sciences technology. This model took into account which nucleotides are preferred within nucleosomes (dinucleotides repeated at ~10-

bp periodicities which accommodate DNA bending) and which 5-mers are preferred within linkers (CGCGC, AAAAA, or A/T 5-mers) (Field et al., 2008). This model successfully predicted the nucleosome occupancy of a single chromosome using a model trained on all other chromosomes.

An important finding in the study by Field et al. was the strong role of nucleosome excluding sequences in positioning nucleosomes. Poly(dA:dT) tracts are one of the strongest nucleosome excluding sequences (Field et al., 2008). They consist of long stretches, 5 to 35-bp, of dAs or dTs that exclude nucleosomes at promoters, origins of replication and gene terminators (Segal and Widom, 2009). Nucleosomes are excluded from both perfect and imperfect poly(dA:dT) tracts allowing proteins access to these sequences (Segal and Widom, 2009). Nucleosome depletion at poly(dA:dT) tracts can be predicted based on DNA sequence alone; this depletion can extend in a window of up to 150-bp surrounding the poly(dA:dT) tract (Segal and Widom, 2009). Transcription factor binding sites near poly(dA:dT) tracts are not the cause of nucleosome depletion because transcription factor binding sites without adjacent poly(dA:dT) tracts are only weakly nucleosome-depleted (Segal and Widom, 2009). Thus, nucleosome-excluding poly(dA:dT) tracts (5 to 35-bp) enhance transcription factor binding site accessibility. One explanation for nucleosome depletion at poly(dA:dT) tracts is their poor affinity for nucleosome formation (Segal and Widom, 2009). Poly(dA:dT) tracts have length-dependent structural properties such as minor groove size, which decreases cooperatively with the length of the tract, resulting in a unique hydration structure with multiple layers of ordered water molecules H-bonding to each other and DNA bases resulting in length-dependent structural properties (Field et al., 2008; Woods et al., 2004). This unique structure requires more energy to be deformed into a nucleosome compared to other sequences (Field et al., 2008). The strong boundary to nucleosome formation created by a poly(dA:dT) tract creates a NDR because there are a smaller

number of nucleosome configurations in which DNA bases are not close to the boundary (Segal and Widom, 2009). The ability of poly(dA:dT) tracts to encode nucleosomes has been shown experimentally (Raisner et al., 2005). Insertion of poly(A) DNA and a Reb1-binding site generated a NDR much larger than the 22-bp of inserted sequence (Raisner et al., 2005). Thus, poly(dA:dT) tracts have a role in specifying nucleosome locations of eukaryotic genomes (Segal and Widom, 2009).

A recent study (Kaplan et al., 2009) has challenged theories which state that nucleosome positioning in yeast is determined through the combined action of chromatin remodellers, DNA-binding proteins, and the DNA sequence preferences of nucleosomes. By generating an *in vitro* nucleosome map of purified histone octamers (from chicken cells) assembled onto purified yeast genomic DNA using salt gradient dialysis, DNA sequence preferences were found to have a substantial influence on nucleosome positioning (Kaplan et al., 2009). *In vitro* nucleosome depletion is found at many transcription factor binding sites, gene start and end sites, reflecting sequence-directed nucleosome depletion (Kaplan et al., 2009). Kaplan et al. measured the average nucleosome occupancy as the number of DNA sequence fragments (reads) over a base compared to the genome-wide coverage per base pair. *In vitro* and *in vivo* nucleosome locations were found to have a correlation coefficient of 0.74 (Kaplan et al., 2009). The similarity between *in vivo* and *in vitro* nucleosome maps indicates the locations of many nucleosomes are not influenced by other DNA binding proteins; instead, nucleosomes appear to have an innate preference for particular genomic locations. Some of the differences in nucleosome locations between *in vivo* and *in vitro* nucleosome maps may be a result of chromatin remodellers moving nucleosomes to less preferred locations, i.e., the 10-bp periodicity of DNA dinucleotides (AT minor groove and GC major groove) which accommodate DNA bending within the nucleosome is less prominent *in vivo* than *in vitro* (Kaplan et al., 2009). A predictive model using

nucleosome sequence preferences from this dataset was designed to distinguish nucleosome-enriched and nucleosome-depleted regions (Kaplan et al., 2009). Three *in vivo* nucleosome maps generated under conditions which cause large-scale transcriptional changes had localized differences and were highly correlated with the *in vitro* nucleosome map (Kaplan et al., 2009). One important difference between the *in vitro* and *in vivo* nucleosome maps was that long-range ordering of nucleosomes is present only *in vivo* but not *in vitro.* On average ChIP-determined transcription factor binding sites were nucleosome depleted *in vivo* and *in vitro,* nucleosome depleted sites had a correlation coefficient of 0.62 between *in vitro* and *in vivo* datasets. Abf1 and Reb1 binding sites were on average more depleted *in vivo* than *in vitro*. This result demonstrated the ability of Abf1 and Reb1 to generate their own nucleosome depletion. Importantly, this study showed that nucleosome depletion around regulatory protein binding sites is largely attributed to DNA sequence, allowing transcription factors increased access to binding sites which contribute to transcription initiation (Kaplan et al., 2009).

The conclusions of Kaplan et al. are in contrast to those from a recent study (Zhang et al., 2009) which determined nucleosome positions from living yeast cells, and nucleosomes assembled onto yeast genomic DNA using purified histones with salt dialysis with or without ACF (a protein that functions in ATP-dependent chromatin assembly). The two studies were not performed identically, Zhang et al. used the *in vivo* 1:1 mass ratio of histones to DNA while the study by Kaplan et al. reported precipitation problems at this ratio and opted for a 2:5 mass ratio of histone to DNA. The lower ratio presumably allowed histones to select optimal DNA sequences (Zhang et al., 2009). The inclusion of the chromatin assembly protein, ACF, during the *in vitro* assembly of nucleosomes generated an *in vivo* linker size of ~20-bp (the *in vitro* linker size is shorter in the absence of ACF) and enhanced the ability to load nucleosomes onto

deproteinized *E. coli* DNA, indicating ACF can position nucleosomes over unfavourable sequences (Zhang et al., 2009).

Zhang et al. determined that translational positioning (variance in the location of sequenced nucleosome midpoints) was lower *in vitro* than *in vivo*. Only ~20% of *in vivo* nucleosome locations are explained by their *in vitro* locations despite the high correlation of *in vitro* and *in vivo* histone densities in both studies: 0.54 (Zhang et al., 2009) and 0.74 (Kaplan et al., 2009). Histone densities do not account for the exact locations of nucleosomes but rather indicate the average histone content per base pair (Zhang et al., 2009). Most differences between *in vivo* and *in vitro* nucleosome positions were at promoters; only a portion of promoters which were nucleosome-depleted *in vivo* were nucleosome-depleted *in vitro* (Zhang et al., 2009). The similarities between *in vivo* and *in vitro* nucleosome positions were at terminators (Zhang et al., 2009).

The study by Zhang et al. provided further insight into the promoter NDR. Encoding the promoter NDR in the DNA sequence, i.e., using poly(dA:dT) tracts, does not assist in the formation of the +1 nucleosome (relative to the transcription start site) because nucleosome positioning is directional, decreasing towards gene terminators (Zhang et al., 2009). The strong positioning of the +1 nucleosome is a result of its positioning relative to transcription initiation and it depends on DNA sequences needed for initiation. It is not clear how to reconcile the results of these two studies and further work on *in vitro* reconstruction is required.

Another study (Field et al., 2009) has investigated the evolutionary importance of sequence-positioned nucleosomes by investigating related yeast species living in different environments: aerobic (*Candida albicans)* or anaerobic (*S. cerevisiae).* Under normal growth conditions, cellular respiration genes are inactive in the anaerobic species while active in the aerobic species

reflecting differences in nucleosome organization at promoters. By measuring promoter nucleosome depletion (using a model which gives the probability per base pair that a sequence is covered by a nucleosome) it was possible to explain divergent expression pattern of genes involved in cellular respiration. Specifically, growth-related genes were found to have open promoters (nucleosome-free) while condition-specific genes have closed promoters (contain a nucleosome) (Field et al., 2009).

Genome-wide nucleosome maps have enhanced our knowledge of transcription and its regulation. For example, it is clear that the locations of nucleosomes are partially sequence determined, and that some nucleosomes are dynamic, repositioned following genetic or physiological perturbation. Other nucleosomal positions can be predicted based on DNA sequence. Finally, the most well positioned nucleosome for coding genes is the +1 nucleosome, which presumably interacts closely with the transcription machinery.

1.2 Yeast origins of replication and the ACS

DNA replication is an essential process needed for cell proliferation. The DNA replication machinery is conserved from *S. cerevisiae* to humans but the sequence motifs that direct the initiation of DNA replication are not (Keich et al., 2008). Replication is initiated from specific sites in the genome, origins of replication. The ~400 origins in *S. cerevisiae* differ in their timing and in the efficiency of origin firing (Knott et al., 2009). As with other DNA transactions, DNA replication occurs within the context of chromatin. In the sections that follow these topics will be described in detail.

1.2.1 DNA replication: an overview of initiation

Cellular viability and proliferation requires the ability to duplicate and segregate genetic information into two daughter cells. Genome duplication involves the initiation of chromosome

replication at specific sites along the chromosome called origins of replication (Huberman and Riggs, 1968). The cell cycle describes the distinct phases of growth, replication and cell division and consists of 4 phases: G_1, S, G_2, and M. During the two gap phases (G_1 and G_2) the cell prepares for DNA synthesis and mitosis through growth by increasing the amounts of proteins and organelles (Rowley et al., 1994). Chromosomes are replicated during S phase and segregated into two daughter cells during M phase. During G_1 phase, if the appropriate extracellular and intracellular conditions are present, the cell becomes committed to DNA replication; this commitment point occurs late in G_1 and is called Start (Hartwell et al., 1974). Proteins required for cell cycle control are conserved across eukaryotes. Many of these proteins have been identified in budding yeast as mutants which arrest at particular points in the cell cycle (Hartwell et al., 1970). Some of the identified proteins have a surveillance role, coordinating distinct cell-cycle events such as chromosome replication and segregation; these proteins are called checkpoint proteins and prevent the cell from progressing to another cell cycle phase before required processes are complete (Rowley et al., 1994). Errors during DNA replication can lead to chromosome loss or deletion or gene loss or mutation (Hartwell, 1992).

DNA replication during S phase begins at hundreds of specific sites in the genome called origins of replication (Raghuraman et al., 2001). Origins are typically intergenic and separated by at least 20-kb (Bell and Dutta, 2002). At origins, two multiprotein complexes called replication forks are assembled. The assembly of the replication fork occurs in a step-wise program. The earliest step, involves the formation of a pre-Replicative Complex (pre-RC) (**Figure 2**). The pre-RC begins to form prior to S phase in the preceding Late M and early G1 phase (Blow and Dutta, 2005). The highly conserved six-subunit origin recognition complex (ORC) initiates pre-RC assembly. In *S. cerevisiae*, ORC binds specific sites within the origin called an ARS consensus sequence (ACS). The Orc1, Orc2, Orc4 and Orc5 subunits, are in close contact with DNA at the

origin, while Orc6 and Orc3 are not (Lee and Bell, 1997). In addition to the ACS, *S. cerevisiae* origins can contain up to 3 B-elements (Marahrens and Stillman, 1992). The B3 element is bound by the transcription factor/chromatin remodelling protein Abf1 (Marahrens and Stillman, 1992). Most origins do not contain a B3 element and instead may be bound by other transcription factors such as Sum1, Rap1, or Mcm1 (Weber et al., 2008). The B1 and B2 elements are easily unwound DNA sequences which may serve as the initial location of DNA unwinding prior to DNA replication initiation (Bell, 1995). ORC interacts with the ACS and the B1 element, a region of ~30-bp, specifically binding to the A-rich strand (Lee and Bell, 1997). The ACS is essential for DNA replication initiation and ORC remains bound to the ACS throughout the cell cycle (Bell and Stillman, 1992).

Pre-RC formation at the ACS (**Figure 2**) is initiated by ORC, which recruits Cdc6 and Cdt1, leading to the recruitment of the mini chromosome maintenance (MCM) helicase at origins (Blow and Dutta, 2005). The abundance of Cdc6 is cell cycle regulated: in early S phase Cdc6 is targeted for degradation following Clb5/Cdc28, cyclin-dependent kinase (CDK), phosphorylation (Elsasser et al., 1999). The cell cycle regulation of Cdc6 levels prevents pre-RC formation outside of G1 phase which could cause re-replication of DNA (Piatti et al., 1996). Cdt1 associates with the C-terminus of Cdc6 at origins to promote MCM protein association with origins (Nishitani et al., 2000). Loading the six-subunit MCM complex (Mcm2-7) is the last step in pre-RC formation. The MCM complex likely functions as a DNA helicase at replication forks (DNA elongation) and origins (DNA replication initiation) (Tye, 1999).

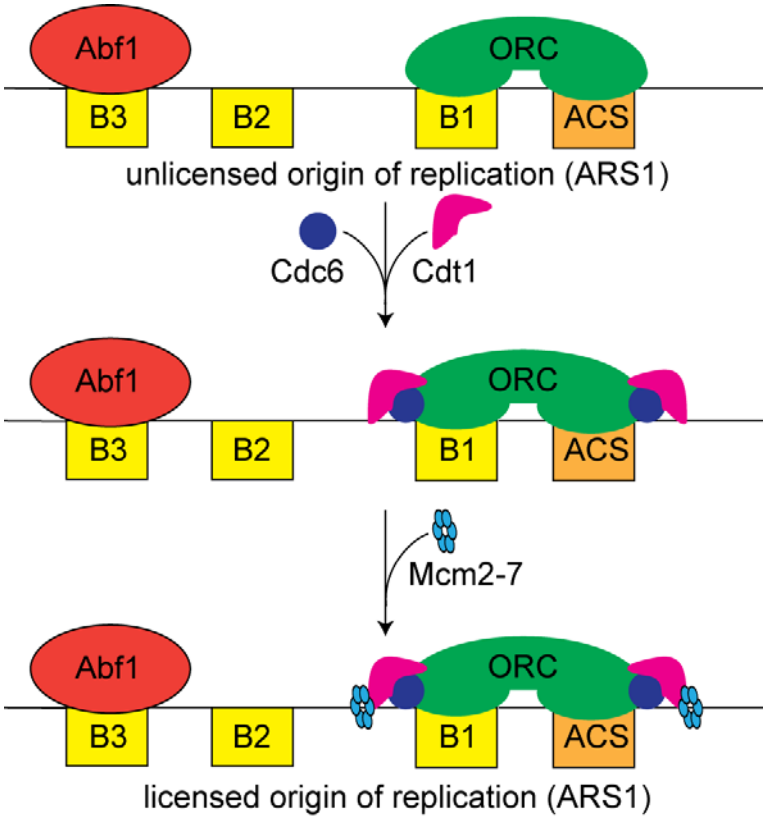

Figure 2: Assembly of the pre-replicative complex leads to an origin licensed for DNA replication.
An origin contains one essential component, the ACS, and as many as three B elements. The six-subunit ORC complex is bound to the ACS and B1 element throughout the cell cycle. The B3 element is present in some origins and is bound by a transcription factor (usually Abf1). Origin licensing occurs between late M and early G1 phase, ORC recruits Cdc6 and Cdt1 leading to the loading of Mcm2-7, the replicative helicase, onto DNA. Once Mcm2-7 is loaded onto DNA, an origin is licensed for DNA replication.

Regulation of pre-RC formation prevents DNA re-replication during the cell cycle. High CDK levels during S phase prevent pre-RC licensing during S, G2 and M phases while allowing origin activation during S phase (Bell and Dutta, 2002). If CDKs containing B-type cyclins (Clb1-6) are inactivated in G2/M using the Clb-Cdc28 inhibitor Sic1 the pre-RC can reform at origins (Dahmann et al., 1995). The genome can be re-replicated from these origins by reactivating CDKs containing B-type cyclins (Dahmann et al., 1995). Cdc28 containing S-phase cyclins

(Clb5 and Clb6) phosphorylate ORC, Cdc6 and MCM to prevent pre-RC licensing outside of G1 phase (Nguyen et al., 2001). Both S phase specific Clb5/Clb6-Cdc28 and G1 phase specific Cln1/Cln2/Cln3-Cdc28 target Cdc6 for degradation (Nguyen et al., 2001). The inappropriate licensing of origins in late G1 and S phases is prevented by several factors, the removal of Cdc6, Orc2 and Orc6 phosphorylation and nuclear export of MCM subunits (Nguyen et al., 2001). This redundancy means that all three of these inhibition mechanisms need to be disrupted for DNA re-replication to occur (Nguyen et al., 2001).

After the cell commits itself to S phase, passing through Start in G1 phase, cyclin B CDKs (Clb-Cdc28) promote the assembly of proteins needed to trigger helicase activation (origin unwinding) and replication fork assembly (Nguyen et al., 2001; Remus and Diffley, 2009). Not all origins where a pre-RC is assembled will fire (MacAlpine and Bell, 2005). In order for DNA synthesis to begin at an origin, several other protein complexes must first associate with the origin (Bell and Dutta, 2002). During the transition from pre-RC to replication forks, Mcm10 may displace Cdt1 from the pre-RC (Bell and Dutta, 2002). Cdc45 and Sld3 are proteins needed for formation of the replication fork. Cdc45 assists with the loading of DNA pol α onto DNA (Aparicio et al., 1999; Mimura and Takisawa, 1998). Once loaded, Cdc45 is a component of the replication fork and helps in the assembly of other fork proteins such as replication protein A (RPA), proliferating cell nuclear antigen (PCNA), GINS complex (Psf1, Psf2, Psf3, Sld5), DNA pol α, δ, and ε (Aparicio et al., 1999; Chesnokov, 2007). Accordingly, the replication timing of an origin correlates with the Cdc45 loading time (Aparicio et al., 1999). DDK (Cdc7 and Dbf4) and CDK (Clb-Cdc28) assist in the transition to DNA replication by phosphorylating replisome proteins (Moldovan et al., 2007). DDK phosphorylates MCM and has a role in recruiting Cdc45 to origins (Bell and Dutta, 2002). CDK phosphorylates Sld2 and Sld3 in order for these proteins

to associate with Dpb11, a required step in fork assembly (Tanaka et al., 2007; Zegerman and Diffley, 2007). Once the replication fork is assembled, replication can proceed.

1.2.2 Origin identification in *S. cerevisiae*

The first *S. cerevisiae* origin to be isolated and characterised was *ARS1* (Stinchcomb et al., 1979). Early methods to identify origins involved fragmenting yeast genomic DNA, inserting these fragments into a vector with a selectable marker, and identifying those fragments which transformed yeast with high efficiency (Stinchcomb et al., 1979). A variety of methods can identify origins in *S. cerevisiae* (Breier et al., 2004; Nieduszynski et al., 2006; Raghuraman et al., 2001; Wyrick et al., 2001; Xu et al., 2006). One approach (Wyrick et al., 2001; Xu et al., 2006) involves cross-linking ORC to its binding sites and, following immunoprecipitation, determining the location of these binding sites by hybridizing the immunoprecipitated DNA to a tiling microarray. This approach can identify origins to within 1-kb (Chesnokov, 2007). Origins can be identified using either sequence conservation within related species (Nieduszynski et al., 2006) or a predictive algorithm (Breier et al., 2004) can be used to identify the functional element within all origins, the ACS which serves as an ORC binding site. Finally, origin identification is possible by determining the locations of newly replicated DNA (Raghuraman et al., 2001; Yabuki et al., 2002) which identified origins at a resolution ranging from 4 to 10-kb (Xu et al., 2006). Origin identification is a necessary step in enhancing our understanding of origin efficiency. For example, it is unclear why only a portion of origins are fired within a population of cell cycles.

Genome-wide location analysis of ORC or MCM binding sites allowed the identification of origins (Wyrick et al., 2001; Xu et al., 2006). These experiments revealed that ~25% of known ARSs were not detectable using ORC ChIP-chip alone, possibly due to differences in the local

chromatin structure (Xu et al., 2006). To precisely locate the ACS within each ARS, a 1-kb window surrounding ORC and/or MCM enriched regions was scanned using an extended position weight matrix (PWM) of the ACS and B1 element based on 31 experimentally confirmed ACSs (Xu et al., 2006). This resulted in the identification of 506 ACSs within 370 potential ARSs (Xu et al., 2006). If the PWM was used to scan the entire genome for ACSs it would have identified 3271 ACSs (Xu et al., 2006). 17 of the ACSs predicted on chromosome 10 were tested using a plasmid-based site-directed mutagenesis approach to remove the essential ACS, showing that 82% of tested ACSs are essential for ARS function (Xu et al., 2006). Caveats of this approach are the small sample size and that the identified origins tend to be efficient (Wyrick et al., 2001).

An alternative approach to identify ACSs involves integrating several data sources: phylogenetic conservation, motif searching and genome-wide location analysis of ORC and MCM (Nieduszynski et al., 2006). Functional origin sequences tend to be conserved among *sensu stricto Saccharomyces* species (Nieduszynski et al., 2005). In order to compile a high quality list of conserved ACS sequences origin locations from several datasets were used: known restriction fragments carrying origins, ORC and Mcm2-7 ChIP-chip enriched regions, and early replicating segments within the genome (Nieduszynski et al., 2006). In this approach, 228 origins containing ACSs were confirmed using a transformation assay which assessed the ability of identified origins to support replication of a plasmid containing a selectable marker (Nieduszynski et al., 2006). Using the precise locations of 228 origins Nieduszynski et al. concluded that origins tend to be located within convergent transcription units and prefer to be closer to transcription terminators.

Using a model (Oriscan) based on the sequence of 26 known ACSs it was possible to identify ACSs genome-wide (Breier et al., 2004). The model incorporated the ACS and flanking regions in the form of a position-weight matrix (PWM). Flanking regions, especially the region 3' to the ACS had a high proportion of A-residues (Breier et al., 2004). The region -108 to +159 around the ACS (described as a PWM) was used to represent 26 known ACSs (Breier et al., 2004). Oriscan analysis consisted of 3 sequential steps: (1) Identification of the top 12,000 matches to the 17-bp ACS PWM; (2) Filtering the list of ACS matches based on the retention of highly conserved positions within the ACS; (3) Filtering the remaining ACSs based on their flanking sequences followed by the rank ordering of all ACS calls (Breier et al., 2004). ACSs were scored based on their proximity to ORC/MCM ChIP-chip defined origins (Wyrick et al., 2001) at 1-kb resolution +/- 250-bp (Breier et al., 2004). Of the top 100 predicted ACSs, 84 correspond to known origins. 10 of the 16 newly predicted ARSs were confirmed using the plasmid assay (Breier et al., 2004). Oriscan did not detect all origins because some origins have more than 4 mismatches to the ACS (Breier et al., 2004).

The methods described in this section have led to the identification of 732 origins (Nieduszynski et al., 2007). Most origins are intergenic and are separated from each other by up to ~100-kb (Diller and Raghuraman, 1994; Raghuraman et al., 2001). Only a fraction of these origins (~228) have experimentally verified ORC binding sites (ACSs) (Nieduszynski et al., 2006). The ChIP-chip defined origins (~370) have multiple potential ACSs per origin, additional studies are required to determine which ACSs are essential (Xu et al., 2006). The Oriscan model identified ~350 origins using the sequence information from a set of 26 well-characterized origins, this training set may have missed real origins while identifying many non-functional origins (Breier et al., 2004). In summary, only a small set of origins (~278) has a verified ORC binding site. In

section **1.2.4** the importance of the ORC binding site in determining the nucleosome positions surrounding origins will be discussed.

1.2.3 DNA replication timing

An important unresolved question regarding origins is why different origins replicate at different times during S phase (i.e., origins have a particular firing time during S-phase) (Raghuraman et al., 2001). In a plasmid, most ARSs replicate early in S phase, while in the context of chromatin, some ARSs are early while others are late (Friedman et al., 1996). The timing of an origin is related to its chromosomal context, for example, moving an early and efficient (used in >90% of cell cycles) origin (*ARS1*) to the subtelomeric location of a late and efficient origin (*ARS501*) converts *ARS1* into a late origin (Diller and Raghuraman, 1994). In addition, there is a tendency for early origins to be near transcribed genes (Diller and Raghuraman, 1994). Replication timing does not affect pre-RC assembly but does have an influence on replication fork assembly (Bell and Dutta, 2002).

One approach to determine DNA replication kinetics involves determining the sites of incorporation of light DNA isotopes within cells containing heavy isotope (^{13}C and ^{15}N) labelled DNA (Fangman et al., 1983). Heavy isotope labelled cells are arrested and released into media containing light isotopes (McCarroll and Fangman, 1988). By collecting samples throughout S-phase and separating light from heavy DNA using cesium chloride density-gradient centrifugation it is possible to distinguish early replicating sequences from late replicating sequences (McCarroll and Fangman, 1988). To identify early and late sequences, density-gradient separated fractions of heavy and light DNA are hybridized to a microarray allowing the percentage of heavy and light DNA to be followed throughout a S phase time-course (Raghuraman et al., 2001). Converting the percentage of heavy/light DNA into replication times

revealed that origins show a continuum of activation times within S phase (Raghuraman et al., 2001). The replication time of centromere-proximal (within 10-kb) origins is earlier than subtelomeric regions. Subtelomeric regions are not always the last sequences to be replicated, for example, a region 280-kb from the left telomere on chromosome 4 is later than most subtelomeric regions (Raghuraman et al., 2001). Nevertheless, origins within ~25-kb of a centromere are significantly (~5min) earlier than an average origin (27.8min) while origins within ~35-kb of a telomere are significantly (~5min) later than an average origin (Raghuraman et al., 2001).

Another approach to determine DNA replication kinetics involves measuring changes in copy number from one to two copies during DNA replication using a microarray (Yabuki et al., 2002). Using flow cytometry the change in relative DNA content following the release of cells from a late G1 block with α-factor was calculated (Yabuki et al., 2002). A replication timing profile was obtained using DNA content values to scale the log2 intensity values obtained following the comparison of each hybridized time point against arrested cells (Yabuki et al., 2002). In contrast to a replication profile based on DNA density (Raghuraman et al., 2001), the copy-number replication profile revealed two origin classes: early and late which differ in terms of their average replication time (Yabuki et al., 2002). These groups corresponded to origins classified as late or early based on their ability to replicate in the presence of the ribonucleotide reductase inhibitor hydroxyurea (HU) (Yabuki et al., 2002) which inhibits origin firing at late origins.

Mapping the genome-wide locations of single-stranded DNA formed in the presence of HU can also reveal the locations of early origins (wild-type) and early/late origins (using a checkpoint deficient *rad53* mutant) (Feng et al., 2006). Treatment with HU causes cells to accumulate single-stranded DNA (Feng et al., 2006). Single-stranded DNA was differentially labelled by

incorporating fluorescent deoxyribonucleotides using random priming and DNA synthesis without denaturation (Feng et al., 2006). Locations with single-stranded DNA are detected by hybridization to a tiling array and correspond to early origins (Feng et al., 2006).

At the level of a single cell, replication timing might be a stochastic process (Czajkowsky et al., 2008). This conclusion was based on results from DNA combing analysis of yeast chromosome 6. Different chromosome 6 fibers (individual chromosome 6 molecules) had different patterns of origin firing (Czajkowsky et al., 2008). Averaging individually distinct patterns of origin firing in 1.25-kb segments smoothed over a 10-kb region generated a replication profile (Czajkowsky et al., 2008) similar to the replication profile generated using density to distinguish newly replicated from unreplicated segments of DNA within a population of cells (Raghuraman et al., 2001). Thus, temporal regulation of origin activation might be a population property rather than representing differences in structure at individual origins. This conclusion is controversial because a mutant (*clb5*) which affects the initiation of origins in early S phase had a significant influence on the replication timing of late-replicating regions of the genome (McCune et al., 2008). The microarray approach in which an entire population of cells in S phase is pooled into a single hybridization cannot be directly compared to a technique in which only a short ~5min pulse of label (DNA combing) is used. The different conclusions of these studies can be reconciled by each origin having a range of times at which it is most likely to fire within an individual cell (McCune et al., 2008). Furthermore, different cells may or may not fire an origin in a particular cell cycle leading to apparent disorder at the level of single chromosome fibers.

1.2.4 Nucleosome organization at origins

Differences in replication timing could result from differences in chromatin structure (Aparicio et al., 2004). Specifically, the accessibility of proteins needed in the initiation of DNA replication

may be influenced by chromatin structure (Vogelauer et al., 2002). Consistent with this, relocating origins to different regions in the genome such as telomeres and silent mating type loci causes a delay in origin replication time (Friedman et al., 1996). Similarly, an origin's late replication timing is maintained on a plasmid only if the plasmid contains enough flanking DNA (~15kb) further suggesting that chromatin architecture influences origin function (Friedman et al., 1996). Several studies involving the chromatin modifying SIR complex have suggested a role for chromatin architecture and replication origins. Sir2 is a histone deacetylase and part of the SIR complex which assembles heterochromatin and delays replication timing at subtelomeric origins (Stevenson and Gottschling, 1999). Delayed replication of subtelomeric origins is lost through the mutation of Sir3, a SIR complex component that binds the tails of histones H3 and H4 (Stevenson and Gottschling, 1999). Origins outside of subtelomeric regions may have their nucleosomes deacetylated by Sir2 (Crampton et al., 2008). These origins contain a sequence element I^S within adjacent nucleosomes which promotes the formation of unfavourable chromatin and inhibits pre-RC assembly (Crampton et al., 2008). All origins are thought to have a pre-RC (ORC, Cdc6, Cdt1) assembled on them during G1 phase. The ability of recruited proteins such as MCM and Cdc45 to bind and activate origins during S phase may be influenced by repressive nucleosome structure (Stevenson and Gottschling, 1999).

Histone deacetylation by Rpd3 has a role in regulating origins not regulated by the SIR complex (Aparicio et al., 2004). Deletion of *RPD3* decreased the replication timing of late origins (non-telomeric) (Aparicio et al., 2004). The earlier replication timing of late origins was accompanied by increased histone acetylation (Aparicio et al., 2004). Targeting a histone acetyltransferase to a late origin causes an earlier replication time (Vogelauer et al., 2002). By measuring the replication timing of all origins within *rpd3*Δ cells, 104 origins were found to be delayed by Rpd3 (Knott et al., 2009). Replication timing was measured using BrdU-IP ChIP, in which

increased BrdU peak height corresponds to earlier initiation and more efficient origin firing (Knott et al., 2009). These authors suggested that histone deacetylation causes chromatin compaction which can delay origin firing (Knott et al., 2009).

In addition to possibly explaining replication timing of different origins, the nucleosome structure of origins plays a role in the assembly of the pre-RC during G1 phase. In order for ORC to be bound to the ACS, the surrounding DNA must be within a nucleosome-free region. Single origin studies confirm this prediction: a nucleosome positioned over the *ARS416/ARS1* ACS inactivates the origin (Simpson, 1990). The positioning of nucleosomes adjacent to the *ARS1* nucleosome free region containing the ACS is influenced by ORC (Lipford and Bell, 2001). Disruption of the nucleosome arrangement adjacent to origins interferes with replication initiation (Lipford and Bell, 2001). Disruption of the ACS leads to nucleosome encroachment into *ARS1* and *ARS307* (Lipford and Bell, 2001). Insertion of sequences which expand the size of the nucleosome-depleted region (e.g. an Abf1 binding sites or a lac operator) on the same side as the ACS resulted in the ACS-proximal nucleosome shifting away from the ACS (Lipford and Bell, 2001). The shift in nucleosome positioning was accompanied by a 3.5-fold increase in plasmid loss rate suggestive of a reduction in origin firing due to an initiation defect (Lipford and Bell, 2001). When the NDR was increased, MCM binding to the origins was reduced and a defect in pre-RC assembly was observed (Lipford and Bell, 2001). Finally, ORC-positioned nucleosomes are necessary for pre-RC assembly.

1.3 Rationale for Thesis

Several studies have examined nucleosome positioning around origins. Chromosome 3 origins were found to be located within nucleosome free regions (Nieduszynski et al., 2006; Yuan et al., 2005). Several other groups (Albert et al., 2007; Field et al., 2008; Mavrich et al., 2008; Yin et

al., 2009) have concluded that origins are on average nucleosome-depleted genome-wide. However, these studies provide average views, and do not investigate the role of nucleosome architecture to explain origin properties. By focusing on a well characterized subset of origins, those with a known ACS, it is possible to infer the nucleosome architecture at origins with a characterized ACS. By determining the nucleosome occupancy at these origins it is possible to determine the consistency of nucleosome positioning at origins. Further, the influence of nucleosome positioning on origin replication times can be determined. Finally, using an inducible ORC mutant the sequence contribution to nucleosome positioning at origins can be investigated, i.e., if origin nucleosomes are sequence encoded, their positioning is not expected to change in the absence of ORC. In summary, defining nucleosome architecture at origins may explain differences in replication timing; further, using appropriate mutants, the impact of ORC on nucleosome positioning at origins can be quantified.

Chapter 2
Materials and Methods

2.1 Nucleosome organization at replication origins

In this section, wild-type refers to a published S288C nucleosomal dataset (Lee et al., 2007). The tiling array coordinates within this dataset refer to a February 2006 genome release from SGD (Hirschman et al., 2006). ACS coordinates (Nieduszynski et al., 2006) for 228 origins refer to an October 2003 release. In order to locate these ACSs within the February 2006 genome, the 15-bp proACS for each origin was used to search the corresponding chromosomal sequence in order to find its location(s). In cases where more than one match was found (N=8 origins), the closest ACS to the described ACS was chosen as the 2006 proACS. A coordinate was assigned to each ACS, as the minimum of its start/end proACS coordinates. Using SGD chromosomal features from February 2006, 65 ACSs were located. SGD proACS calls are 11-bp long; to locate the 15-bp proACS, the minimum of ACS start/end sites were subtracted by 2. These ACSs were annotated with their ORIdb identifier, and the entire list of Nieduszynski et al. and SGD ACSs were filtered for duplicate calls. This resulted in a list of 278 ACS calls (228 Nieduszynski + 50 SGD). This list was then filtered based on the criteria that at least 800-bp of flanking sequence (the window size used to analyze origins) was located on either side of the ACS (255 ACSs).

ACS proximal probes, all probes within 800-bp of the ACS were localized and made into a text file where each position 0 represents the nearest ACS probe. When a probe is not located within a 4-bp window, the value was assigned as NA. The orientation of the ACS, which strand (Watson or Crick) is the T-rich strand of the ACS, was taken into account by flipping the entire list of extracted (-)-sense, T-rich strand on the Crick strand, log2 values. This list was imported into the software program R, and scaled so that each origin-proximal region has a mean of 0 and

standard deviation of 1. The sequence of steps needed to obtain the log2 values surrounding the ACS are summarized in a flowchart (**Figure 3**).

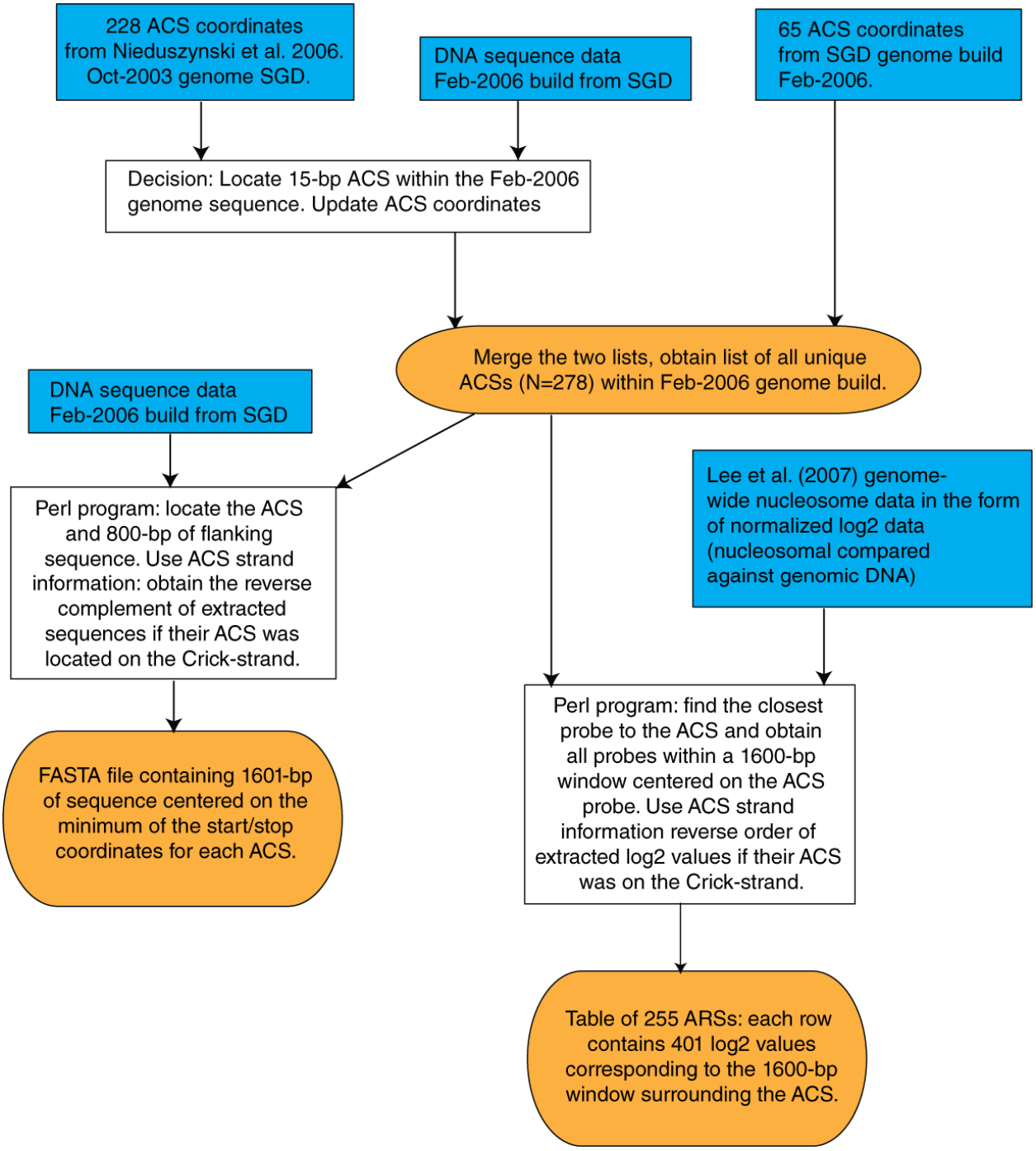

Figure 3: Flowchart describing the process to obtain ACS-centered origin sequence and ACS-centered nucleosome profiles.

Using R (R Development Core Team, 2009) the mean-ACS centered ACS profile was generated and overlaid onto a bivariate histogram (**Figure 8**), generated using the hexbin package (Carr et al., 2009). The hexbin serves as a two-dimensional error bar for each point within the mean ACS profile. As a comparison, a random subset of coding genes was obtained using a random number generator (Eddelbuettel, 2009) to pick 255 genes from a list of 5015 coding genes (Lee et al., 2007). To calculate the average size of nucleosome NDRs in ARSs and coding gene profiles, the locations of nucleosome midpoints, peak log2 values, were visually selected using R and the distance between points was printed onto the figure (**Figure 9**).

2.2 Nucleosome occupancy at replication origins correlates with dinucleotide sequence features

A list of 103 DNA dinucleotide properties were obtained from the DiProDB website (Friedel et al., 2009). The sequence of 255 oriented origins was used to count dinucleotides within 75-bp windows using the count function of the Seqinr package (Charif and Lobry, 2007). At each window, the dinucleotide counts were multiplied by the corresponding property value, summed for all dinucleotides and divided by the total number of dinucleotides in the window. This value was then assigned to the central probe. In order to determine correlation with the wild-type nucleosome profile, the average dinucleotide property at each position was calculated, and compared to corresponding log2 probes using Pearson correlation. The process used to correlate DNA dinucleotide properties with the nucleosome occupancy at origins is summarized in a flowchart (**Figure 4**).

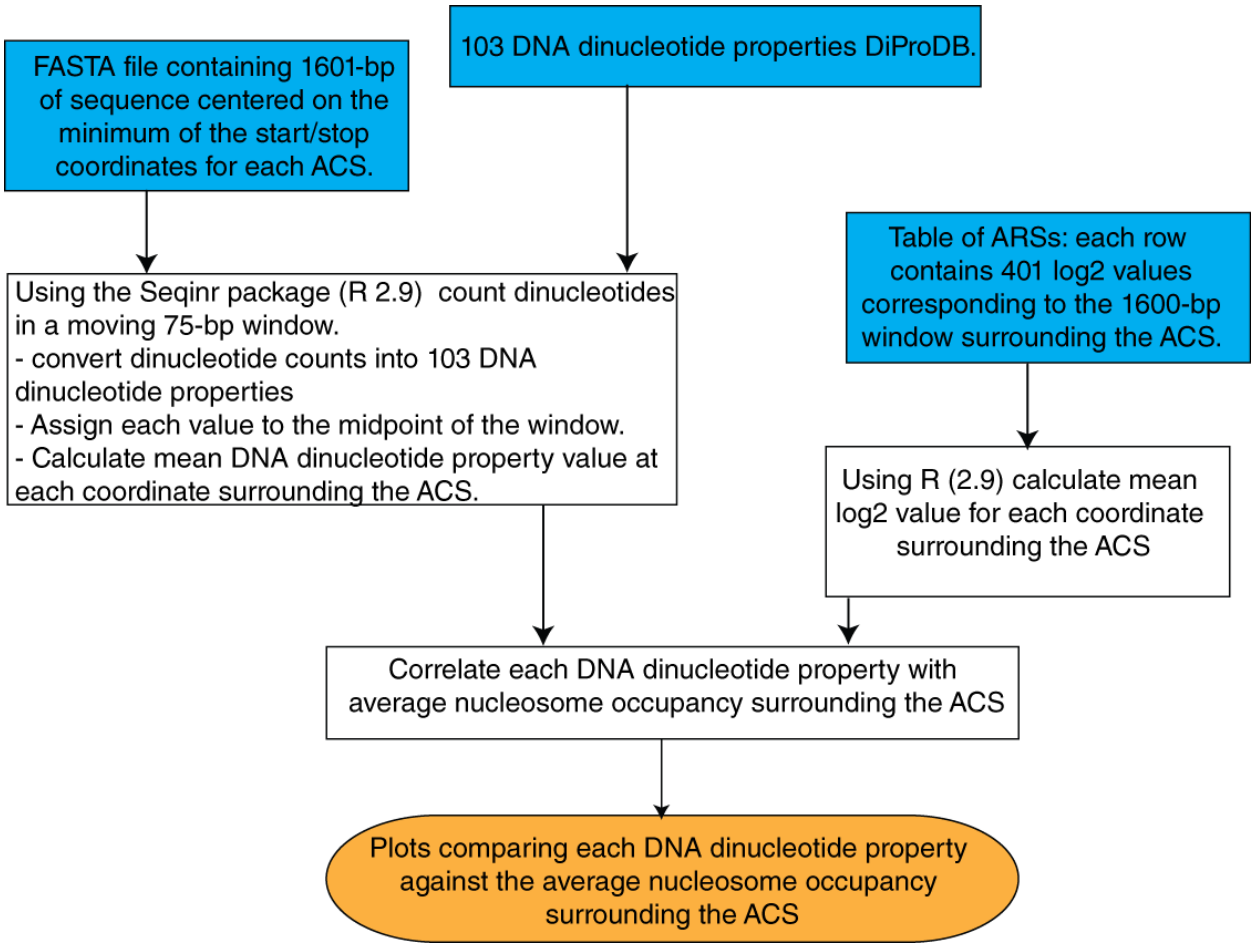

Figure 4: Flowchart describing the process to obtain plots comparing DNA dinucleotide properties with ACS-centered nucleosome profiles.

2.3 Clustering analysis reveals distinct nucleosome occupancy signatures at replication origins

The analysis performed in this section is presented as a flowchart in **Figure 5**. The 800-bp region centered on the ACS was clustered using Ward's method and the R-implementation of agnes hierarchical clustering (Maechler et al., 2005). The dissimilarity matrix for clustering was obtained using uncentered Pearson correlation calculated using the amap package (Lucas, 2009). The resulting dendrogram was cut using the dynamicTreeCut package (Langfelder et al., 2008) with parameters deepSplit set at 3 and minimum cluster size set at 20. Detecting clusters in a dendrogram involves cutting branches off the dendrogram. The dynamicTreeCut package is a

hybrid of hierarchical clustering and partitioning around medoids. This algorithm does not rely on using a standard cut height: branches are cut based on their shape. In the first stage of this analysis, clusters must contain a minimum number of objects (I chose N=20 after testing an array of values), outliers within the same branch are removed from a cluster if their distance is too far from other members of the cluster, and clusters must be distinct from surroundings. In the second stage, the dendrogram is ignored and dissimilarity information is used to assign unassigned objects to a cluster using a method similar to partitioning around medoids. The heatmap was constructed using the heatmap.2 function of the gplots package (Warnes et al., 2009). Subclustered nucleosome occupancy signatures were constructed by averaging only those origins within a cluster. The extent of the NDR was calculated by visually locating peaks, and using R to calculate the distance between the closest data points.

An extended proACS motif was obtained by extracting the region -10 to +40 around the ACS start site, position 0. This sequence was used as input for the command-line version of weblogo 3.0 (Crooks et al., 2004), which took into account the background base frequencies of *S. cerevisiae*. Abf1 binding sites within an 800-bp region of the ACS were identified by scanning ACS-aligned sequences in a moving window of 16-bp, width of the Abf1 position weight matrix (PWM). Each 16-mer was assigned a PWM score by looking up Abf1 PWM values for each position and summing the values together. A PWM is a motif representation of a DNA-binding protein's specificity (MacIsaac and Fraenkel, 2006). The PWM motif is represented in the form of a matrix where the width of the matrix corresponds to the motif length and each column corresponds to a position in the motif which contains the probability of observing a particular base at that position (MacIsaac and Fraenkel, 2006). PWM motifs are often visualized using a sequence logo where the height of letters at each position represents the information content which ranges from 0 (each nucleotide has an equal probability of occurring) to 2 bits (one base is

always found) and the relative heights of letters indicate the probability of observing a particular base (MacIsaac and Fraenkel, 2006). The cut-off for detecting Abf1 binding sites involved identifying Abf1 binding sites in all coding genes and selecting the top 250 unique PWM scores (Lee et al., 2007). Values greater than the cut-off were counted for each origin using a moving window of 20-bp.

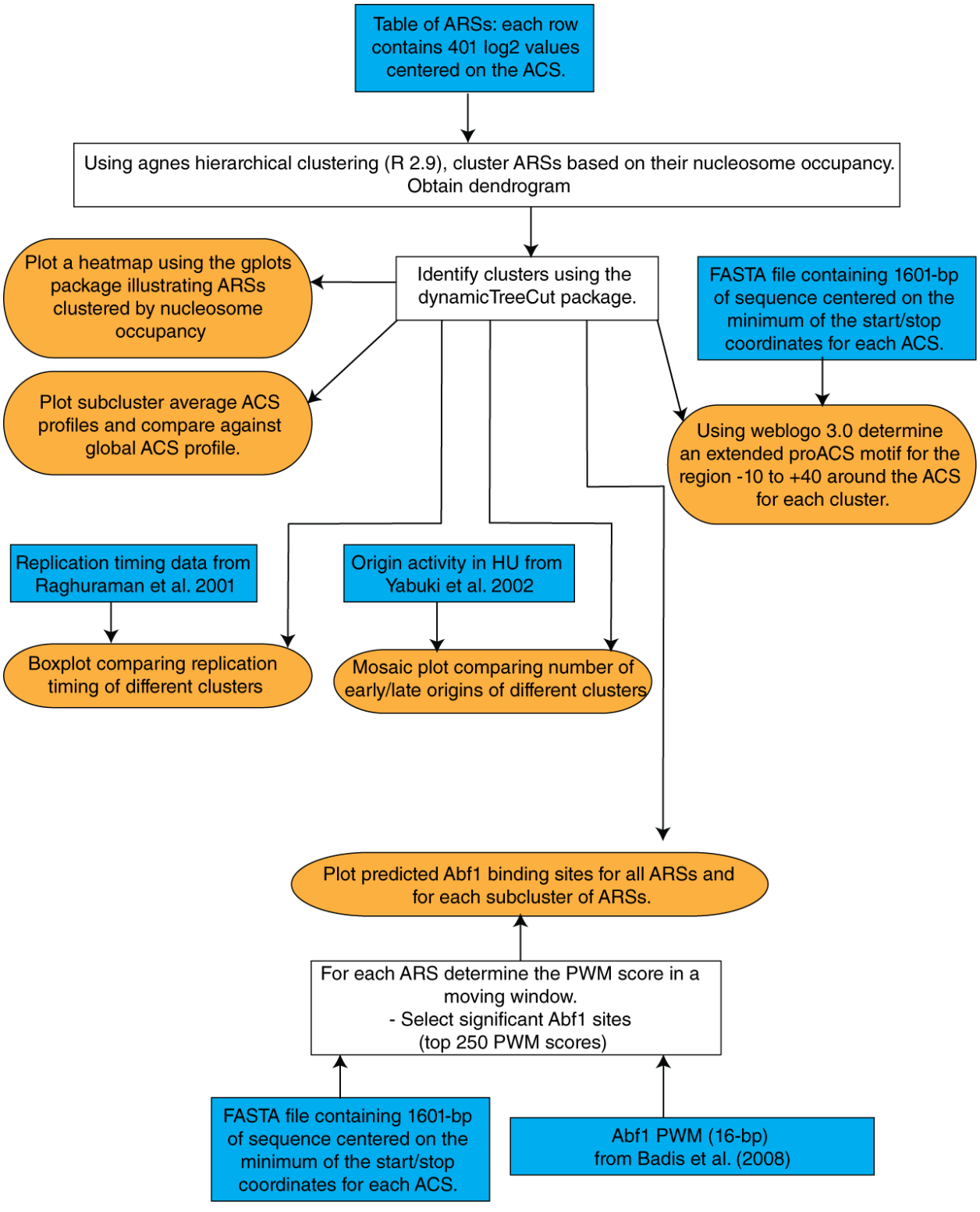

Figure 5: Flowchart describing the analysis of wild-type nucleosome profiles.

2.4 Nucleosome occupancy signatures correlate with origin activity in hydroxyurea

Replication timing (Raghuraman et al., 2001; Yabuki et al., 2002) as well as origin activity in HU (Feng et al., 2006) was obtained from OriDB (Nieduszynski et al., 2007). Replication timing data for the subset of origins with identified ACSs Yabuki et al., (N=181) and Raghuraman et al. (N=185) were grouped according to their clustering groups and analyzed using an analysis of variance test to determine if there were any significant differences between mean cluster replication time.

In contrast to the replication timing data, more origins have activity in HU data (N=254). The replication timing data for origins was grouped according to their origin nucleosome signature and tabulated. Using a chi-square test, it was possible to determine if there was an association between origin nucleosome signature and origin activity in HU. The cross-tabulation data is displayed using a mosaic plot, from the vcd package (Meyer et al., 2009). To identify which clusters were responsible for the association of origin nucleosome signatures with replication timing each cluster was compared to its expected number of early and late origins. Expected values correspond to the proportional number of early and late origins. Using a chi-square test for each cluster, groups with significant differences in the number of early/late origins were identified.

The genomic context of each origin (N=255) was determined by comparing the location of the ACS against a list of genomic features: coding gene start/end sites (http://chemogenomics.stanford.edu/supplements/03nuc/files/clusters/polyA_segments_verified_coords.txt), telomeres/centromeres (http://downloads.yeastgenome.org/chromosomal_feature/archive/SGD_features.tab.200602.gz)

and the locations of all ARSs (http://www.oridb.org) localized to the February 2006 genome release using BLAT (http://genome-test.cse.ucsc.edu/~kent/exe/).

Table 1: Strain List

Strain	Genotype
W303-1A	*MATα ade2-1 trp1-1 his3-11,15 ura3-1 leu2-3,112 can1-100*
GAL:orc2-1	*MATα ade2-1 trp1-1 his3-11,15 ura3-1 leu2-3,112 can1-100 orc2-1::Pgal1-3HA-orc2-1/TRP1*
BY4741	*MATa his3Δ0 ura3Δ0 leu2Δ0 met15Δ0*

2.5 Binding of the origin recognition complex positions nucleosomes at origins

The microarray (PN 520055) used in this study contains the double stranded *S. cerevisiae* genome tiled with probes offset by 4-bp on average (Lee et al., 2007). The protocol used to obtain nucleosomal DNA via micrococcal nuclease digestion is described elsewhere (Lee et al., 2007). Changes to this protocol include increasing the size of the yeast cultures from 50mL to 200mL. Single colonies of either W303-1A (the wild-type strain) or *GAL:orc2-1* (Shimada et al., 2002) were inoculated into 25mL of YPAG (1% yeast extract, 2% tryptone, 0.04% adenine sulphate, 2% galactose) and grown overnight (~20h) at 30°C. The cultures were diluted to an OD ~ 0.1 in a final volume of 200mL YPAG in a baffled 1L flask. Cultures were grown until an OD ~ 0.6 (~1 x 10^7cell /mL) and then blocked with nocodazole (Sigma) at a final concentration of 5µg/mL with 1% DMSO. Cells were blocked for 90 minutes, collected and resuspended in 200mL YPAD containing 5µg/mL Nocodazole and 1% DMSO. Cells were blocked in YPAD for 60 minutes, collected and released into 200mL YPAD. Time points were collected every 15 minutes from 30 minutes to 2 hours after the release from a nocodazole block and analyzed by FACS (Davierwala et al., 2005). The sample at the final time point, 2 hours, was cross-linked using methanol-free formaldehyde at a final concentration of 2% for 30 minutes. After the

formaldehyde was quenched using 125mM glycine for 5 minutes, the cells were collected in a 250mL centrifuge tube, washed with 1X PBS and collected into a 50mL Falcon tube. The cell pellet containing ~4 x 10^9 cells was frozen using liquid N_2 and stored at -80°C.

Nucleosomes were isolated from 200mL of cross-linked cells (~4 x 10^9) by digesting the cell wall using zymolyase (Seikugaku 20T) at a final concentration of 0.5mg/mL with 24mL of Zymolyase buffer [1M Sorbitol; 50mM Tris pH 7.4; 10mM β-mercaptoethanol] for 30-45 minutes at 30°C with rotation. Spheroplasting was monitored by taking a small sample (100µL) of the zymolyase reaction diluted 1 in10 into a cuvette, and monitoring the decrease in OD over time. The OD of zymolyased cells begins at ~10 and decreases to ~0.5 within 30 minutes. Cells were collected at 5000xg for 10 minutes and resuspended in 10mL MNase buffer [2 ml of 1M Sorbitol; 50 mM NaCl; 10 mMTris (pH 7.4); 5 mM MgCl2; 1 mM CaCl2 and 0.075% NP40, with freshly added 1 mM β-mercaptoethanol and 500 mM spermidine]. Micrococcal nuclease (Worthington) 7.18 Units/mL was prepared by adding 9mL of molecular grade water (Sigma) directly to the MNase powder, the MNase solution was aliquoted into PCR tubes and frozen at -20°C. Micrococcal nuclease was added in a gradient from 0 to 9µL in 1µL increments to 1mL of spheroplasted and crosslinked cells. The 0µL MNase sample served as a genomic DNA control. The reactions were incubated for 30 min in a 37°C water bath and stopped using 125µL of stop buffer [5% SDS; 100mM EDTA] and 5µL of 20mg/mL Proteinase K (Fermentas) followed by a 16-20h reversal of crosslinks at 65°C. DNA was isolated using a phenol-extraction, followed by a phenol-chloroform extraction, followed by ethanol precipitation and resuspension in 50µL of dH2O and 4µL RNase A. RNA was digested for 3h at 37°C followed by ethanol precipitation and resuspension in 45µL H2O. The quality of DNA was assessed using either 2% w/v agarose gels or the Bioanalyzer to quantify the amount of mononucleosomal DNA

(Agilent, Foster City, CA). Microarray labelling and hybridization is described elsewhere (Lee et al., 2007).

Two biological replicates of *GAL:orc2-1* and W303-1A nucleosomal DNA microarrays were obtained along with one biological replicate of W303-1A genomic DNA (http://www.ebi.ac.uk/microarray-as/ae/ **Accession Number**: E-MEXP-2369). To get a view of nucleosome positioning within *GAL:orc2-1* or W303-1A the nucleosomal DNA CEL files were compared against the CEL file of W303-1A genomic DNA using CEL file processing described elsewhere (Lee et al., 2007). To obtain a view of nucleosome occupancy changes between wild-type and *GAL:orc2-1* the two W303-1A CEL files (controls) were compared against the two *GAL:orc2-1* CEL files (treatment) using Affymetrix Tiling Analysis Software using parameters described elsewhere (Lee et al., 2007). The text files from TAS were parsed in a similar manner as the Lee et al., wild-type data: the 1600-bp window-centered on the ACS was extracted and oriented based on which strand contained the T-rich ACS sequence. To highlight differences between *GAL:orc2-1* and W303-1A origins, the text file obtained by comparing nucleosomal arrays of *GAL:orc2-1* vs. W303-1A were analyzed. For each origin the mean of log2 values was calculated on coordinates within a 400-bp region centered on the ACS. These values were clustered using Ward's method of hierarchical clustering with a Euclidean dissimilarity matrix. A heatmap was constructed in a manner analogous to the wild-type nucleosome signature analysis. The sequence of steps used to perform analysis on *GAL:orc2-1* nucleosome profiles are presented as a flowchart (**Figure 6**).

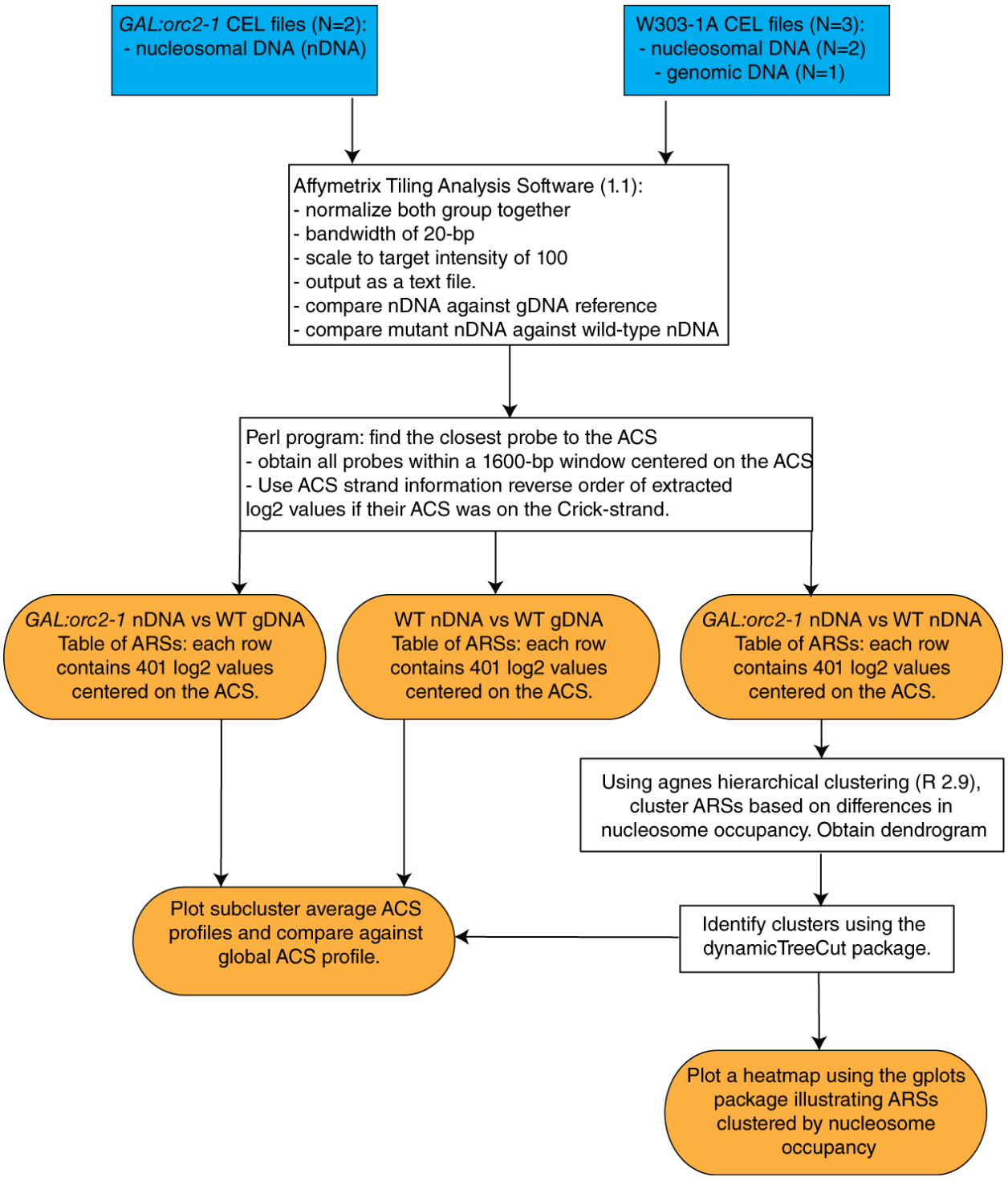

Figure 6: Flowchart describing the process to compare *GAL:orc2-1* and wild-type nucleosome occupancy at origins.

2.6 The ACS remains nucleosome-free when chromatin is assembled *in vitro*

The normalized genome-wide locations of nucleosomes assembled onto deproteinized yeast genomic DNA were obtained (Kaplan et al., 2009). The data file was parsed to obtain the normalized log2 value of the 1600-bp surrounding the ACS start coordinate. This dataset has more missing values compared to the tiling array data. Thus, origins which had at least 75% of coordinates in the 100-bp region surrounding the ACS were used to construct an average ACS profile of *in vitro* nucleosomes. This corresponded to 198 origins. The *in vitro* data was plotted as a bivariate histogram using the same method used to make the wild-type bivariate histogram. The average size of the NDR was calculated by measuring the distance from the two maxima on either side of the NDR.

Websites:

[1] Local sources of SGD sequence data (Feb-2006).
http://hugheslab.ccbr.utoronto.ca/supplementary-data/tillo/nucleosomes/

[2] Lee, W. et al. (2007) wild-type data
http://chemogenomics.stanford.edu/supplements/03nuc/files/analyzed_data_complete_bw20.txt

[3] Description of the S288C genome chip
http://www-sequence.stanford.edu:16080/S288C/

[4] SGD chromosomal features table
http://downloads.yeastgenome.org/chromosomal_feature/archive/SGD_features.tab.200602.gz

[5] Yeast replication origin database (OriDB)
http://www.oridb.org

[6] Microarray data: http://www.ebi.ac.uk/microarray-as/ae/
Accession Number: E-MEXP-2369

Chapter 3
Results

3.1 Nucleosome organization at replication origins

Several groups have investigated the nucleosome occupancy patterns of coding genes (Field et al., 2008; Lee et al., 2007; Mavrich et al., 2008; Shivaswamy et al., 2008). These studies agree on the nucleosomes architecture at coding genes in which an array of nucleosomes extends in the direction of the ORF away from the promoter. The first and most well-positioned nucleosome, the +1 nucleosome, is adjacent to the transcription start site (Lee et al., 2007; Yuan et al., 2005). Limited work has been done towards understanding the nucleosome occupancy at origins (Field et al., 2008; Mavrich et al., 2008; Yin et al., 2009); however, current studies are incomplete and have not aligned origins with respect to the ACS, the ORC-binding site. Aligning with respect to the ACS (**Figure 7**), the ORC binding site, is significant because nucleosomes have been shown to be positioned by ORC (Lipford and Bell, 2001). Previous studies have aligned origins with respect to origin start and end sites, which are usually not functional elements of the origin, but rather are often arbitrarily defined by the location of restriction enzyme cut sites. Previous nucleosome maps using origin start sites lead to the conclusion that origins are within a nucleosome-free region (Yin et al., 2009), but failed to provide any evidence of nucleosome phasing adjacent to the ACS.

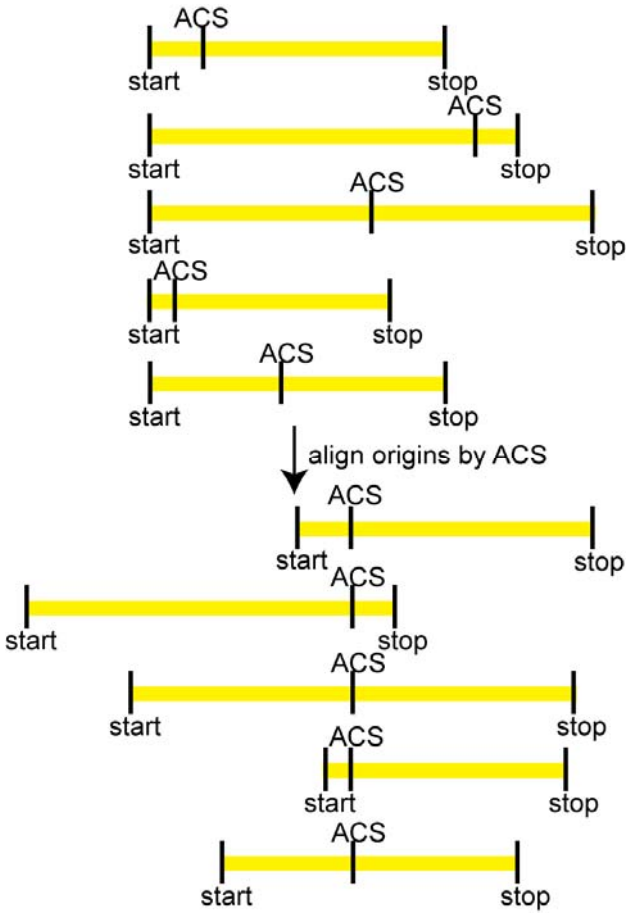

Figure 7: Alignment of origins by the ACS as opposed to origin start sites.
Origins can be aligned using origin start sites (a non-functional origin element) or the ACS (the ORC-binding site).

The ACS-centered view of 255 origins and a random subset of 255 transcription start site-centered coding genes were compared (**Figure 8**)**.** The average view indicates that nucleosomes are well-positioned on either the side of the nucleosome-free region containing the ACS (**Figure 8B)**. The positioning of origin adjacent nucleosomes is comparable to the positioning of the +1 nucleosome within a random subset of coding genes (**Figure 8A)**. In array-based nucleosome calls, an array of nucleosomes is represented by a periodic curve in which local maxima correspond to the midpoint of a nucleosome while minima correspond to a linker region. The amplitude of this curve represents the strength of nucleosome positioning. The ARS nucleosome array extends at least 3 nucleosomes away from the ACS nucleosome-free region, while the

coding gene nucleosome array extends at least 5 nucleosomes away from the promoter NDR. In contrast to directional promoters the nucleosome positioning on either side of the ACS is comparable, i.e., symmetric. The average size of the origin NDR (262-bp) is smaller than the promoter NDR (281-bp) as shown in **Figure 9**. The linker between the ±1 and ±2 nucleosomes is larger in origins than it is in coding genes. The bivariate histogram of origin nucleosome structure (**Figure 8B)** indicates significant variation of individual ACS-centered nucleosome profiles.

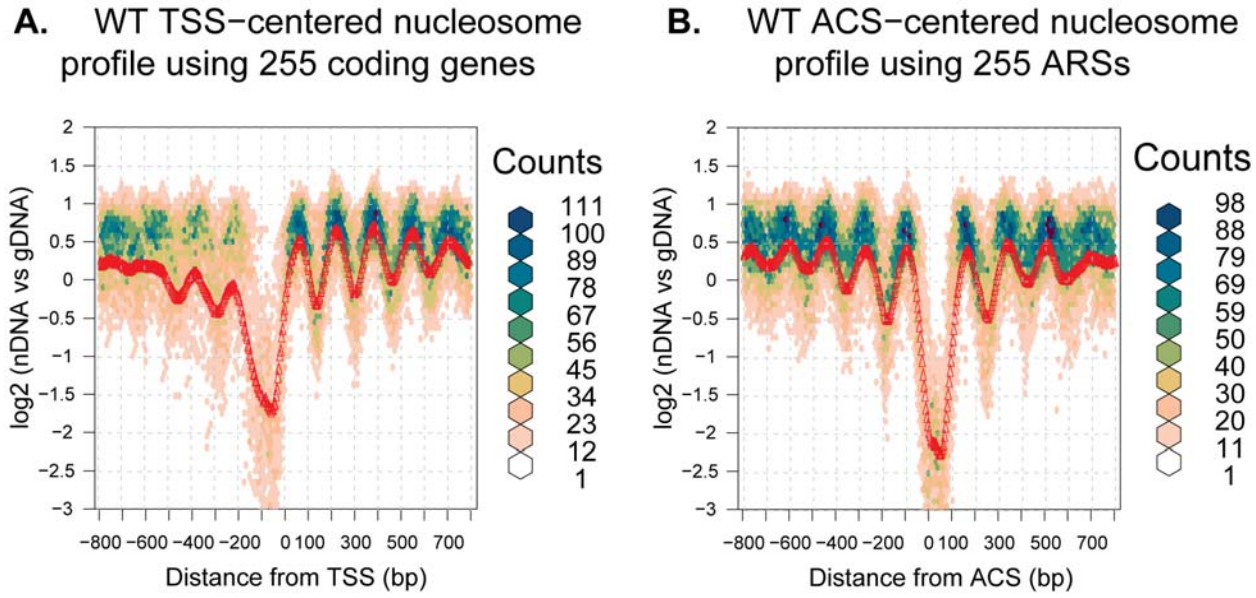

Figure 8: Comparison of transcription start site centered ORFs and ACS-centered ARSs. The diversity within transcription start site (TSS-) or ACS-centered data is represented using a bivariate histogram which represents the density of data within a hexagonal bin as a colour. The distance from the ACS corresponds to the start of the ACS for origins which had their T-rich strand on the Watson strand and the end of the ACS for origins which had their T-rich strand on the Crick strand. Overlaid on this distribution (in red) is the mean TSS- or ACS-centered nucleosome profile. Nucleosome arrays are represented by a periodic curve in which peaks correspond to nucleosome midpoints while troughs correspond to linkers between nucleosomes.

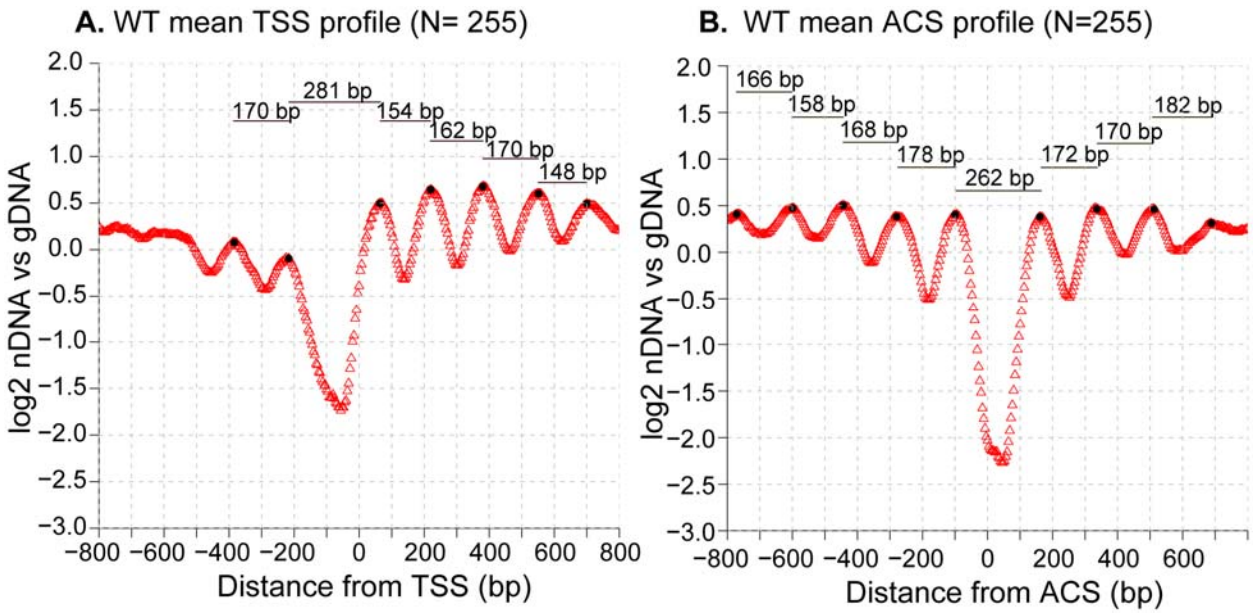

Figure 9: Parameters of nucleosome occupancy at transcription start sites and origins. The distance between adjacent nucleosome midpoints is shown above each nucleosome profile. The size of the coding gene nucleosome-depleted region (NDR) (**A**) is larger than the origin NDR (**B**). The peak-to-peak nucleosome distances of coding genes are smaller than the peak-to-peak nucleosome distances of origins.

3.2 Nucleosome occupancy at replication origins correlates with dinucleotide sequence features

DNA sequence makes a strong contribution to the genome-wide location of nucleosomes (Kaplan et al., 2009; Zhang et al., 2009). Based on nucleosome sequence preferences, it is possible to predict whether or not a particular stretch of DNA is located within a nucleosome (Kaplan et al., 2009). Factors which contribute to nucleosome occupancy at promoters include DNA dinucleotide properties (Lee et al., 2007). The ACS lies within poly(dA:dT) tracts which tend to form an extended NDR (Field et al., 2008). The NDR surrounding the ACS is illustrated by calculating the average GC-content of ACS-centered origins (**Figure 10)**. The average GC-content of origins is highly correlated with the average ACS-centered nucleosome profile, but is unable to explain the locations of nucleosomes because it lacks periodicity. To determine if any DNA dinucleotide properties explained the location of nucleosomes, an exhaustive list of 103

DNA dinucleotide properties (Friedel et al., 2009) was used. The correlation coefficient of each DNA dinucleotide property with the average nucleosome profile was determined (**Figure 11**). Four classes of DNA dinucleotides were identified: (1) High correlation with the origin nucleosome profile, but lacking periodicity to explain nucleosome occupancy (**Figure 12A**); (2) Moderate correlation with origin nucleosome profile and ability to explain nucleosome occupancy to the left of the ACS (**Figure 12B**); (3) Moderate correlation with the origin nucleosome profile predicting a larger NDR (**Figure 12C**); (4) Poor correlation with the origin nucleosome profile (**Figure 12D**). DNA sequence features make a significant contribution to origin nucleosome occupancy patterns, but most features are only able to explain the NDR not the locations of positioned nucleosomes.

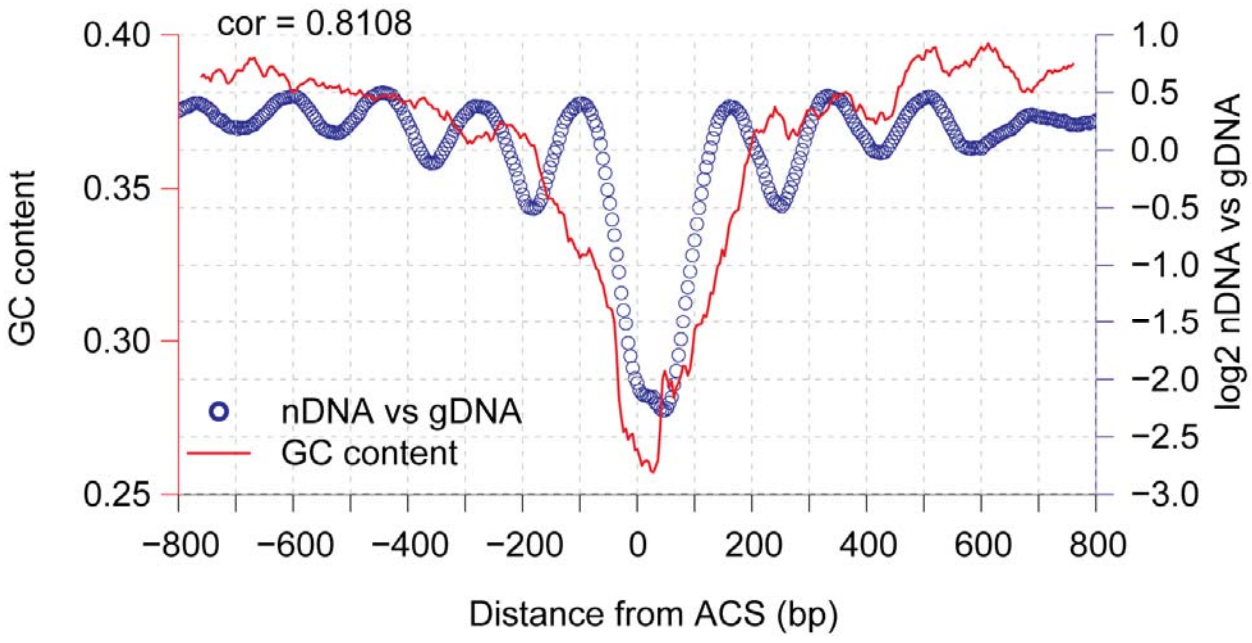

Figure 10: Average GC-content and average ACS-centered nucleosome profile.
The average GC-content of 255 ACS-centered origins was calculated in a 75-bp window. The GC-content was compared against the average ACS-centered nucleosome profile. The ACS lies within an extended NDR. Based on the correlation the periodicity of nucleosomes is not fully explained by GC-content.

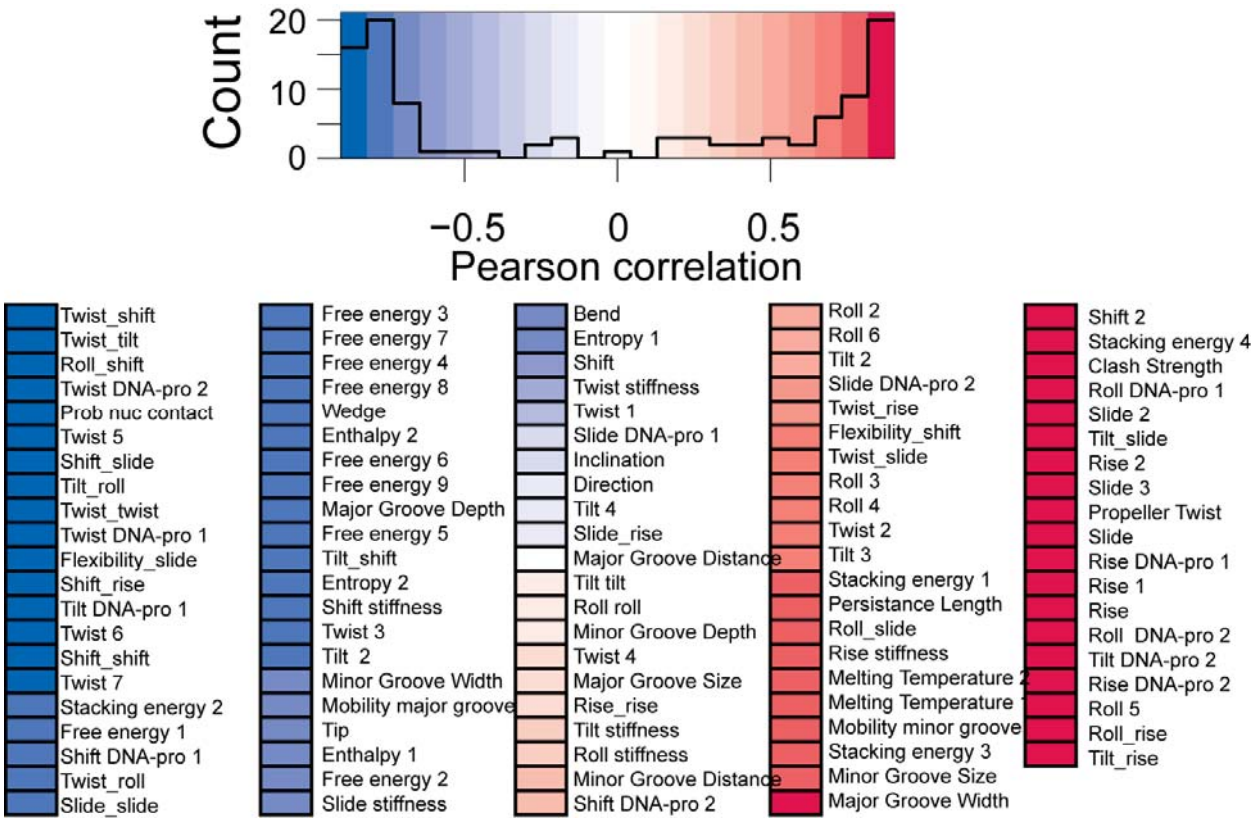

Figure 11: DNA dinucleotide correlation with average origin nucleosome profile.
The correlation of each DNA dinucleotide property (N=103) with the average origin nucleosome profile is shown. The average of each DNA dinucleotide property was calculated in a 75-bp moving window. Generally, most dinucleotide properties are able to explain the NDR surrounding the ACS.

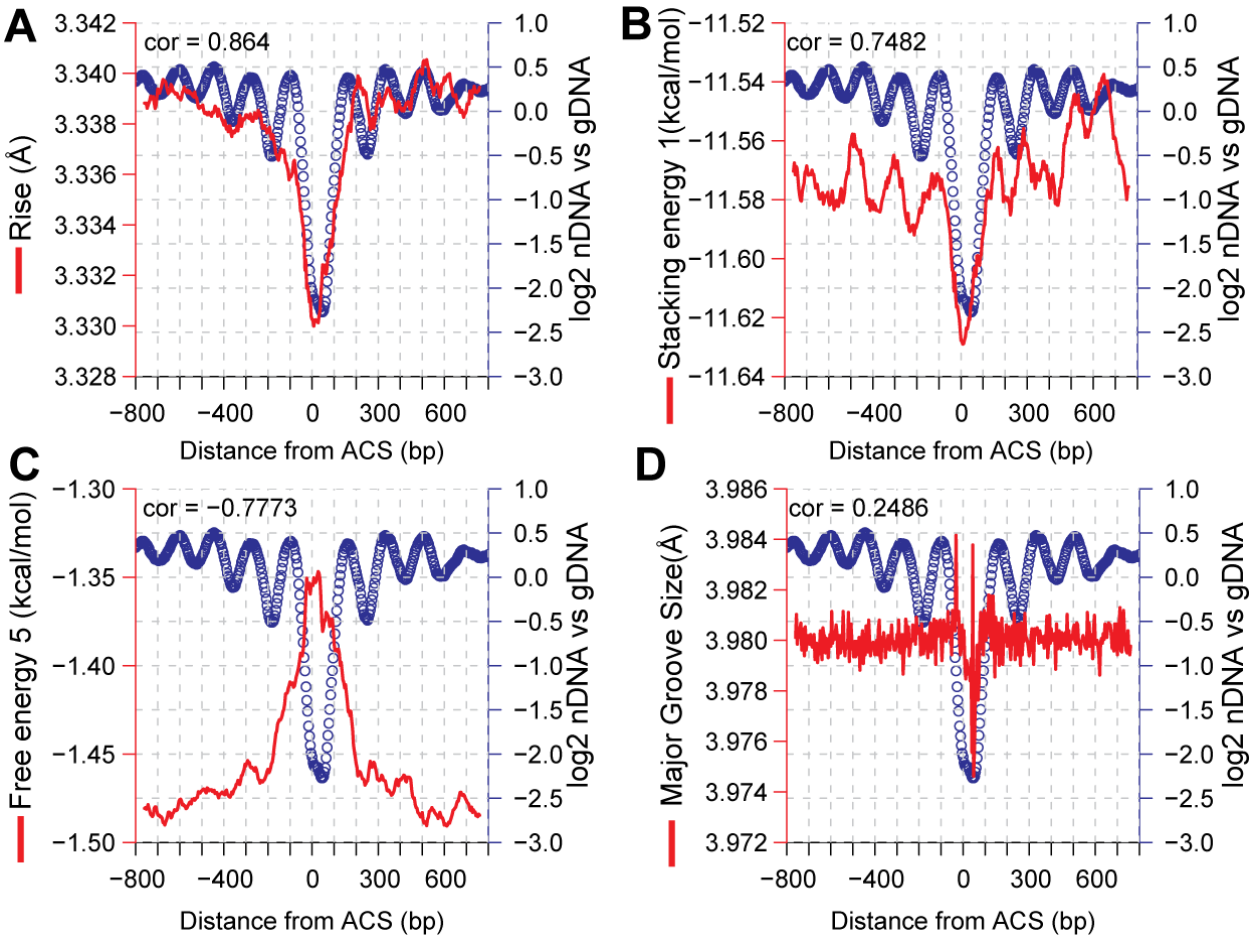

Figure 12: A sample of DNA dinucleotide profiles.
A. The average DNA rise has a high correlation with the average origin nucleosome profile but lacks periodicity to explain nucleosome positioning. **B.** The average stacking energy has moderate correlation with the average nucleosome profile and explains some of the positioning of nucleosomes to the left of the ACS. **C.** The average free energy has moderate correlation with the average nucleosome profile but predicts a more extensive NDR. **D.** Average major groove size has poor correlation with the average nucleosome profile.

3.3 Clustering analysis reveals distinct nucleosome occupancy signatures at replication origins

Differences in chromatin structure may explain differences in origin activity *in vivo*. Hierarchical clustering was used to highlight differences between origins (**Figure 13**). Eight clusters were identified in an unbiased manner (Langfelder et al., 2008) by selecting branches with at least 20 origins followed by the expansion of clusters using between origin dissimilarity information. In general, the ACS ± 50-bp serves as the left border of the NDR which extends ~100-bp to the

right of the ACS. Positioned nucleosomes are located to the left and right of the NDR. Using subcluster averages it is easier to spot deviations between the average and subcluster view of nucleosomes at origins (**Figure 14**). Cluster 1 (green) has a distinct nucleosome profile. There is no extended NDR at the ACS, and nucleosomes are not aligned between origins. Cluster 2, 3 and 4 have similar nucleosome occupancy to the average nucleosome profile. Clusters 5 and half of cluster 6 have a second NDR to the right of the NDR containing the ACS. Half of cluster 7 has a second NDR to the left of the ACS, with two nucleosomes in between the ACS-containing NDR and the second NDR. Cluster 8 has a second NDR to the right of the ACS, with only one nucleosome in between the ACS-containing NDR and the second NDR.

The groups identified using hierarchical clustering will be used to investigate biological differences between clusters. Using a different clustering approach (k-means clustering) it is possible to detect similar nucleosome profiles. K-means clustering arbitrarily selects the number of clusters to partition origins into. In **Figure 15** nucleosome profiles are partitioned into 2 to 5 groups. Distinct nucleosome occupancy patterns become apparent when selecting 5 or more clusters using k-means clustering (**Figure 15D**). In **Figure 15D**, the five classes of origins include: two profiles (turquoise and brown) with a second NDR to the left of the ACS-containing NDR, one profile (blue) with a larger linker between the +1 and +2 nucleosomes, one profile (yellow) which matches the average ACS profile and a profile (green) which lacks both positioned nucleosomes and a NDR. In **Table 2**, the origins within the k-means cluster (K=5) are compared to the origins within the 8 clusters defined using hierarchical clustering. There are some differences in the results obtained by the two clustering methods. Nevertheless, different clustering methods reveal the diversity of nucleosome signatures at replication origins.

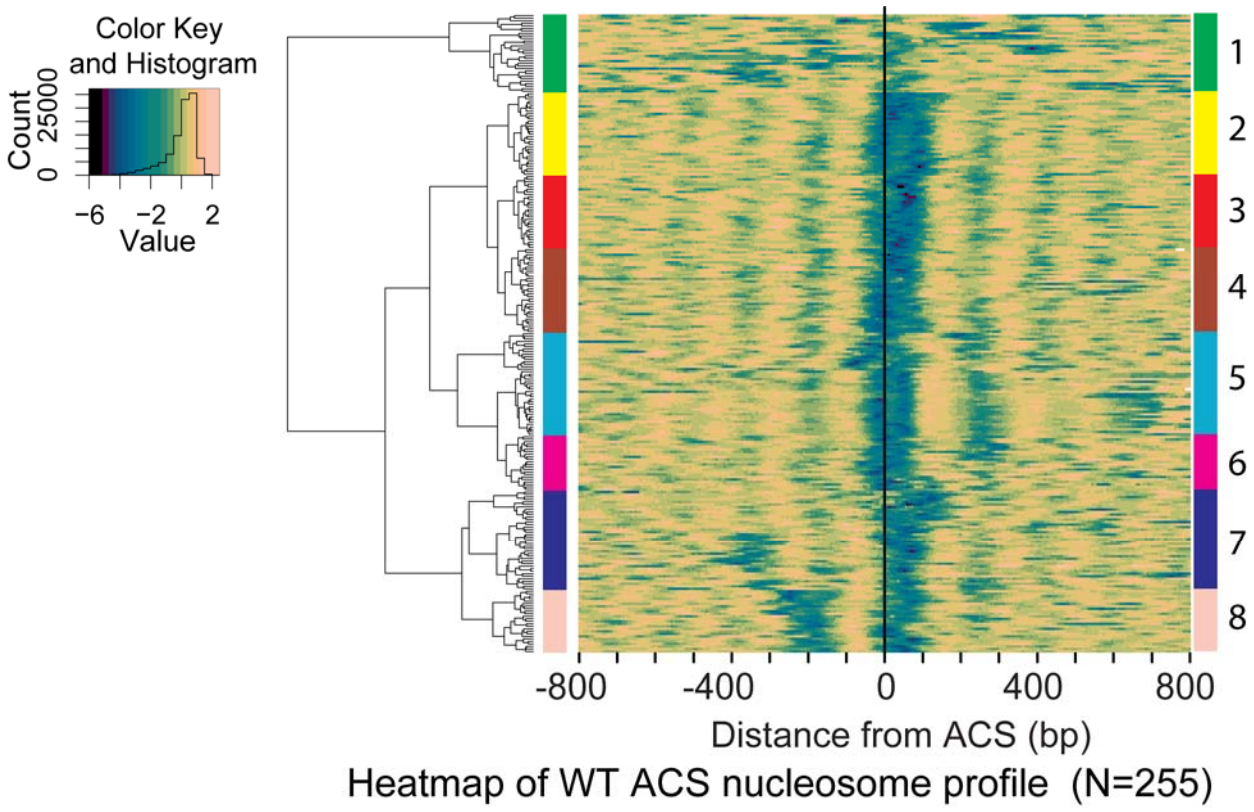

Figure 13: Heatmap of hierarchically clustered, ACS-centered, nucleosome profiles. The log2 values surrounding the ACS (-400 to +400-bp) for each origin were correlated against each other and hierarchically clustered. Distance from the ACS corresponds to the start of ACSs if their T-rich strand is on the Watson strand or end of the ACS if their T-rich strand is on the Crick strand. The resulting dendrogram was used to order a heat map representation of nucleosome occupancy surround the origin. The dendrogram was used to identify groups which illustrate some of the diversity of origin nucleosome profiles. See the main text for a discussion of the differences between the 8 identified clusters.

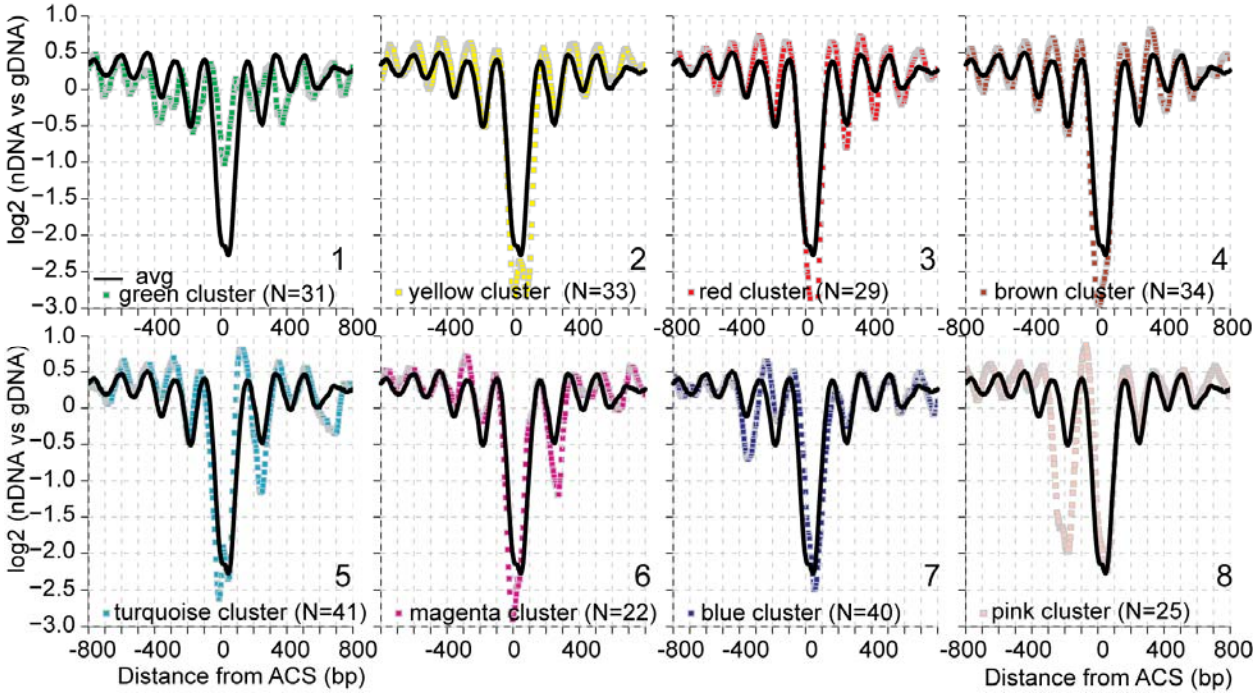

Figure 14: Subcluster average view of clustered origin nucleosome profiles.
Subcluster averages are shown for each cluster identified by hierarchical clustering (Figure 6). In each figure, the average ACS profile is shown in black in order to highlight differences between Individual origin nucleosome profiles. See the main text for a discussion of the differences identified.

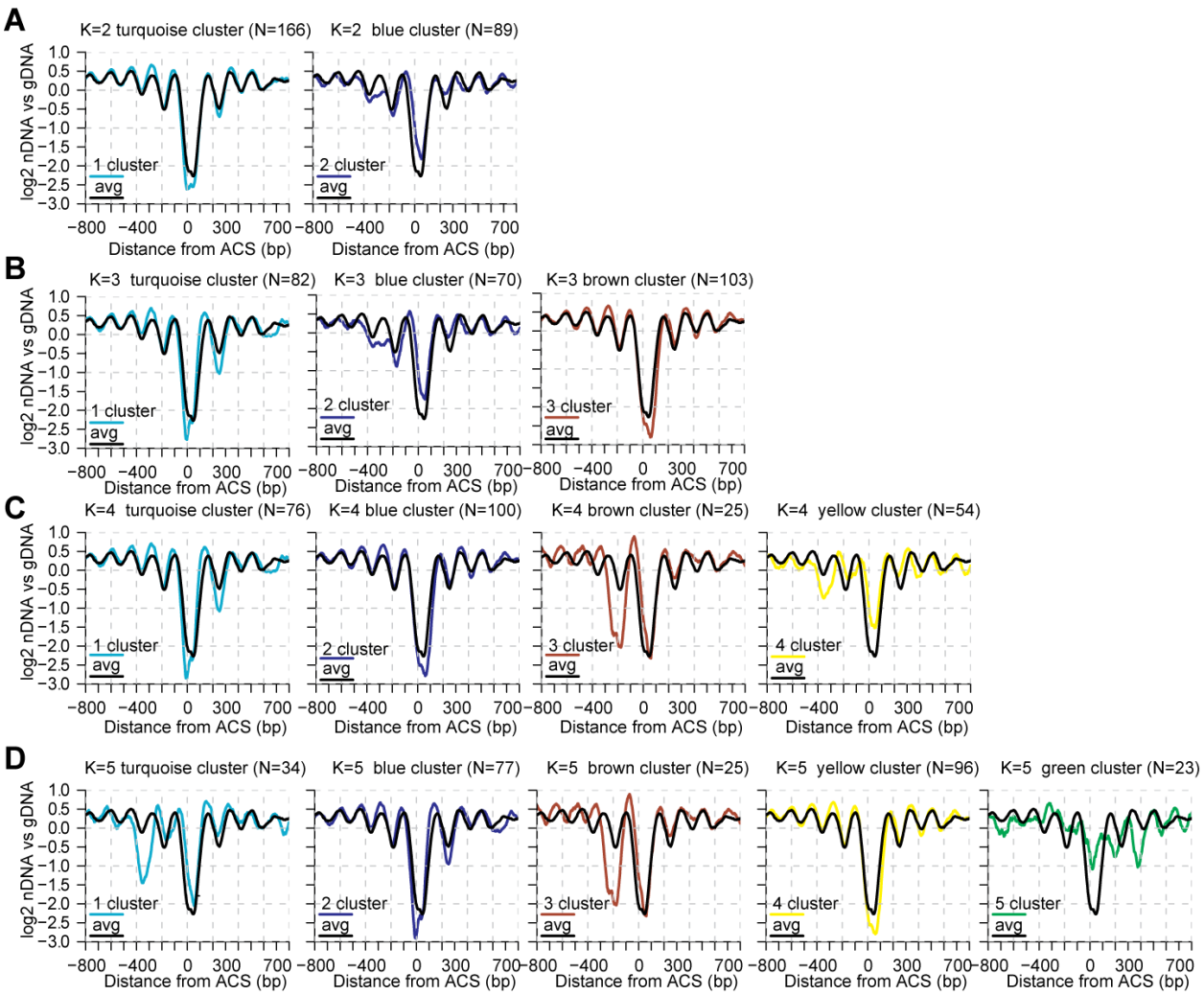

Figure 15: Subcluster average nucleosome occupancy profiles obtained using k-means clustering.

Nucleosome profiles were hierarchically clustered using k-means clustering with 100,000 iterations. The number of clusters was varied between K=2 and K=5. The average profile of each subcluster is shown. Setting the number of clusters to K=5 reveals several distinct nucleosome architectures.

Table 2: Comparison of cluster membership between k-means clustering (K=5) and hierarchical clustering.

		K-means clustering (K=5) defined clusters				
		turquoise	blue	brown	yellow	green
Hierarchical clustering defined clusters	green	12	1	2	0	16
	yellow	0	0	0	33	0
	red	0	0	0	29	0
	brown	1	19	0	14	0
	turquoise	0	37	0	0	4
	magenta	0	18	0	4	0
	blue	21	0	0	16	3
	pink	0	2	23	0	0

Using ACS-aligned sequences it was possible to determine if differences in nucleosome occupancy at origins reflect differences in the ACS and/or adjacent DNA sequences. Differences were detected by identifying motifs in the form of a position weight matrix (PWM) logo (**Figure 16**). To the left of the ACS there was very little information content, each base occurred with approximately equal probability (~0 bits). The highest information content was observed within the 15-bp ACS for all subclusters. The ACS sequence had minor deviations between clusters (**Figure 13**, **Figure 14**): varying in the information content of particular positions. The turquoise cluster in particular had more information content throughout the ACS, indicating most ACSs had a similar sequence. To the right of the ACS, the B1 region was identified as 3-bp with increased information content. The turquoise cluster had higher information content throughout this region indicating the presence of more repetitive DNA, implying the origins were located within telomere-proximal DNA. To investigate this possibility and to determine which chromosomal features were closest to each subcluster the average distance of each cluster of origins to the nearest genomic feature (telomere, centromere, origin and coding gene) was calculated and displayed in the form of a boxplot (**Figure 17**). On average, cluster 5 (turquoise) is very close to telomeres compared to other clusters (**Figure 17A**). Cluster 8 (pink) which had

two adjacent NDRs (**Figure 14**) was the closest to transcription start sites. The closest origins to transcription terminators (**Figure 17B**) were in Cluster 2, which had a nucleosome profile similar to the average ACS nucleosome profile. Cluster 1 (green), which had a unique nucleosome profile (**Figure 14**), was closer to other origins than any other cluster. There were no major differences in the distance of each cluster of origins and their distance to the centromere (**Figure 17E**). Out of the 5 chromosomal features tested, the most significant differences between clusters were: distance of origins to telomeres or gene start and end sites.

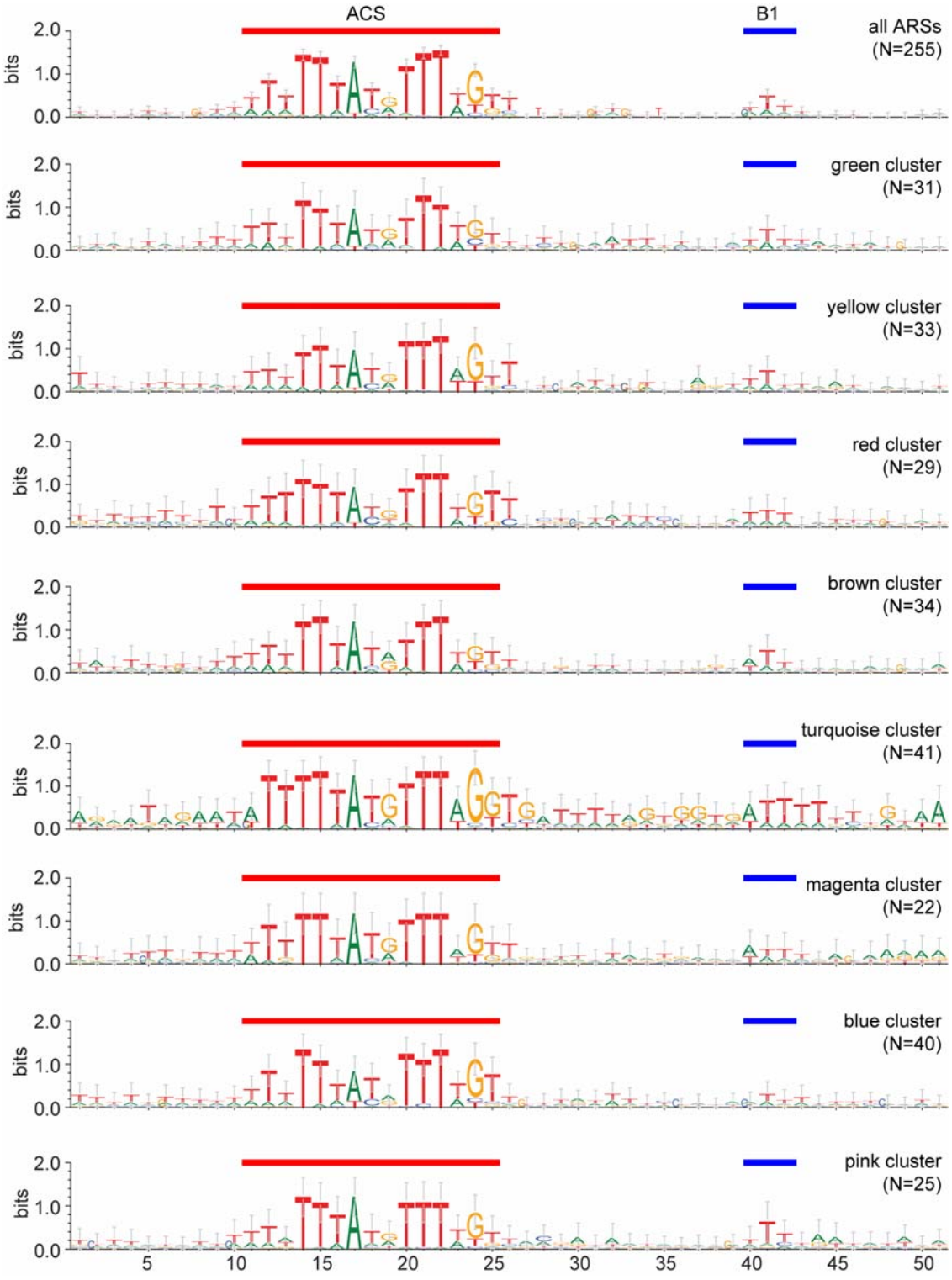

Figure 16: PWM logo of ACS and adjacent sequences.
The sequence logo for all ARSs and each subcluster was constructed using the program WebLogo. The 10-bp upstream of the ACS and the 40-bp downstream of the ACS was examined for any bases with increased information content (bits). A position that is highly conserved will have high information content. See main text for details.

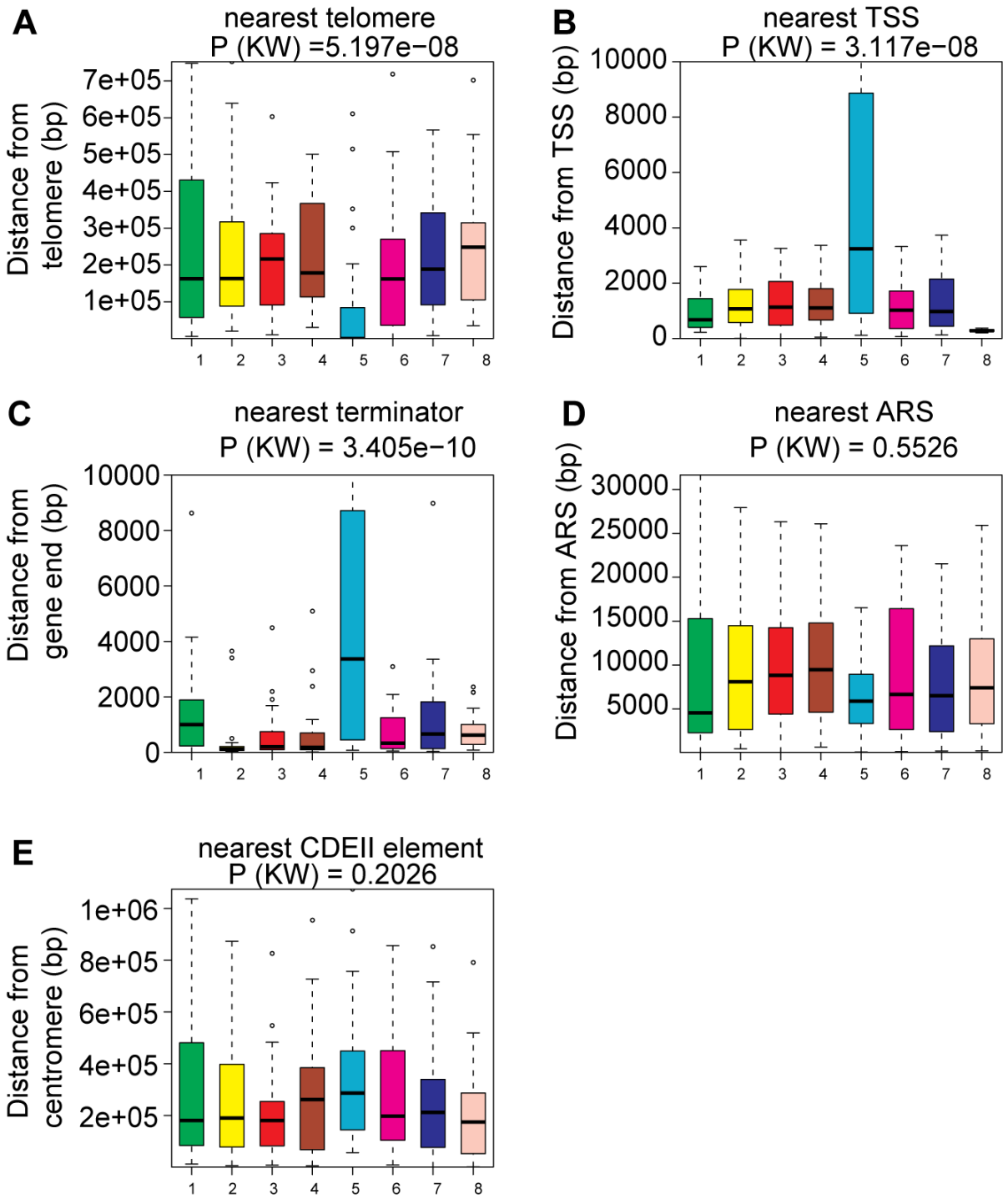

Figure 17: The proximity of each origin subcluster to diverse chromosomal features. The distance of each origin to the nearest chromosomal feature: telomere (A), transcription start site (B), terminator (C), ARS (D), and centromere (E) was calculated and aggregated together based on cluster membership. Each boxplot represents the interquartile range from the first quartile to the third quartile. The whiskers extend either to the minimum or maximum value unless these values are beyond 1.5 times the interquartile range; outliers are represented with circles.

The transcription factor Abf1 has a role in establishing chromatin structure at promoters and origins (Badis et al., 2008; Lipford and Bell, 2001). The top 250 Abf1 PWM scores (Abf1 binding sites) tend to occur within the promoter, 100-bp to the left of the transcription start site (TSS) **Figure 18A** (Lee et al., 2007). In origins, the top 250 Abf1 PWM scores are found ~230-bp to the right of the ACS within the linker separating the +1 and +2 nucleosomes (**Figure 18B**). Sorting origins by their nucleosome profile allows the visualization of Abf1 binding sites within each cluster (**Figure 19**). The turquoise cluster contains most of the Abf1 binding sites. The location of the Abf1 binding site is coincident with the second NDR to the right of the ACS-containing NDR (**Figure 14**). The identification of Abf1 binding sites within this cluster is consistent with telomeric origins sharing a common structure in which the ACS is bordered by an Abf1 binding site (Louis, 1995).

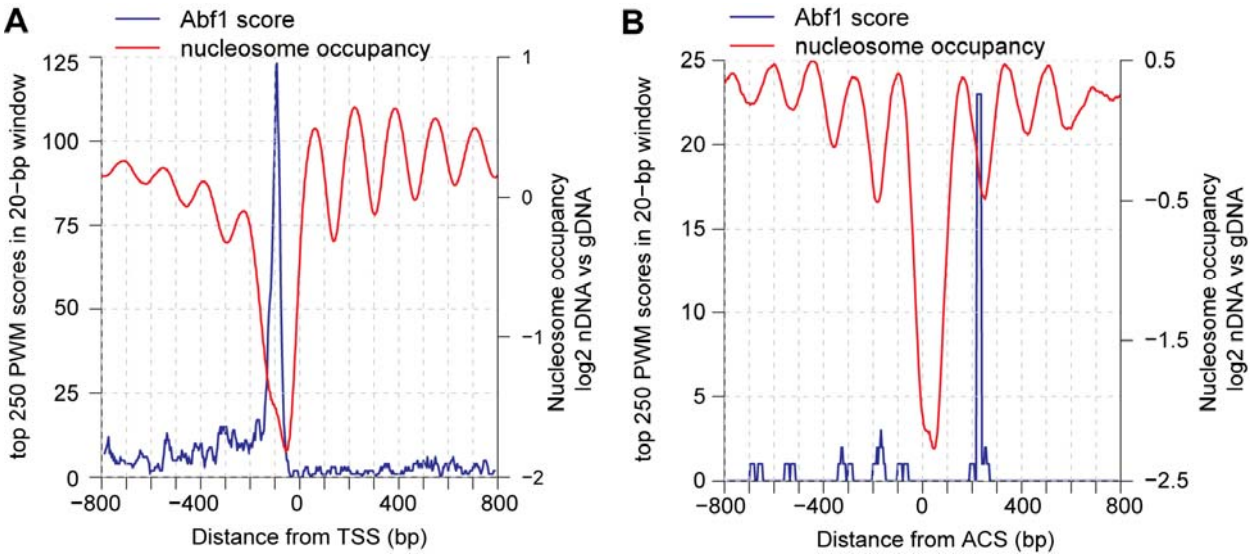

Figure 18: Location of high affinity Abf1 binding sites in coding genes and origins.
Abf1 binding sites are represented in a 16-bp position weight matrix (PWM) (Badis et al., 2008). The sequence of each transcription start site (TSS)-centered coding gene (A) or ACS-centered origin (B) was scored using the Abf1 PWM. The locations of the top 250 Abf1 sites were determined in a moving window of 20-bp and compared against the average nucleosome occupancy for promoters or origins.

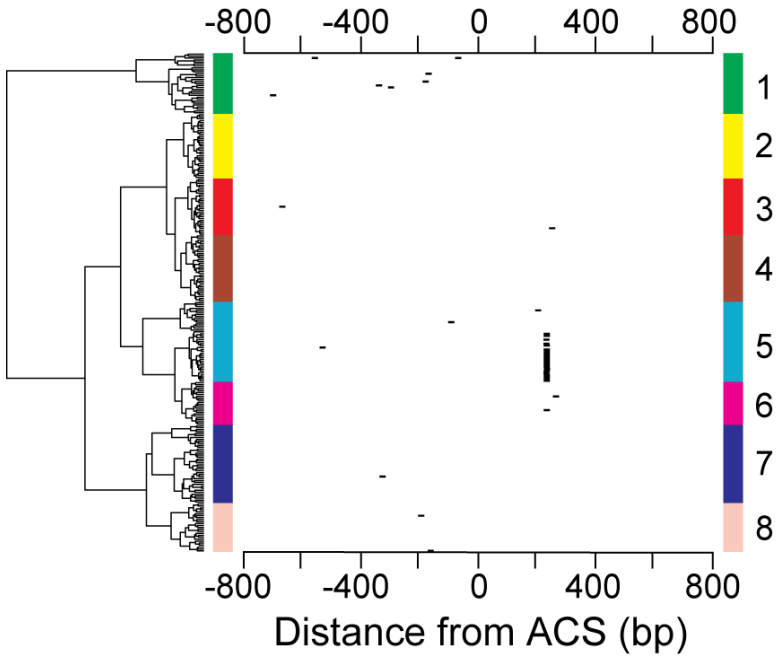

Figure 19: Abf1 binding sites for each origin.
The top 250 Abf1 PWM scores were used to identify Abf1 binding sites within the 1600-bp region surrounding the ACS. Abf1 binding sites were counted in a window of 20-bp for each origin. Individual origins were ordered by the dendrogram obtained by hierarchical clustering (**Figure 13**).

3.4 Nucleosome occupancy signatures correlate with origin activity in hydroxyurea

I tested the hypothesis that differences in chromatin structure might explain differences in origin replication timing. By identifying 8 subclusters it was possible to categorize some of the differences in chromatin structure. Genome-wide replication timing data is available as replication timing profiles for most origins (Raghuraman et al., 2001) or a list of origins which fire in the presence of HU (Feng et al., 2006). Replication timing profiles from ORIdb provide a replication time for only 185 origins (**Figure 20B**). In order to compare the replication timing of all origins, replication timing profiles (Raghuraman et al., 2001) were examined for the local minimum replication time within 5-kb of their ACS; providing data for all 255 origins (**Figure 20A**). Using this revised definition 173 of 185 ORIdb origins had an identical replication time.

The other 12 origins differed up to ~2.3 min between the two datasets. The cluster containing most of the subtelomeric origins (cluster 5) had the latest replication timing. Other clusters varied in their replication times but the differences were not significant.

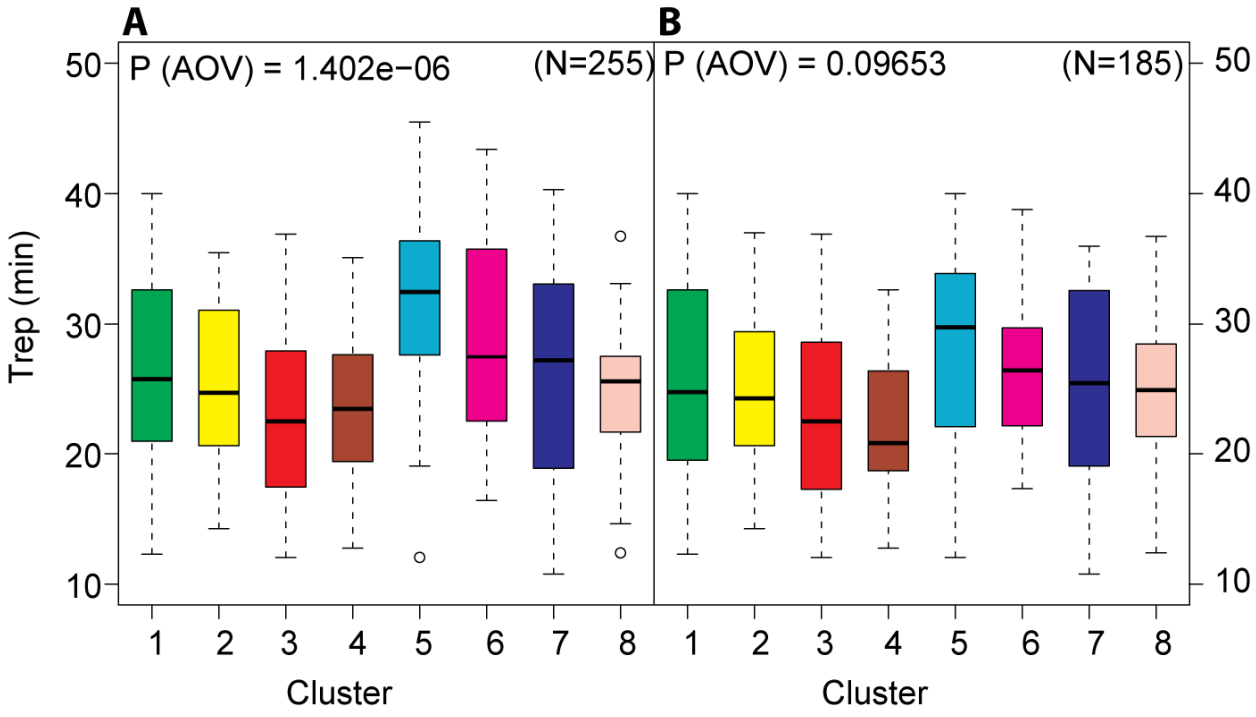

Figure 20: Comparison of average replication timing between clustered nucleosome profiles.
The replication timing (Raghuraman et al., 2001) of each ACS-centered origin was assigned based on the local (10-kb window around the ACS) minimum replication timing value (**A**) or assigned by ORIdb (**B**). When the entire list of origins was used, the average Trep of each cluster were significantly different using an ANOVA test.

Another measure of origin replication time is the ability of an origin to fire in the presence of hydroxyurea (HU) which leads to a block in early S phase. The proportion of early (active in HU) and late (inactive in HU) origins within each subcluster was determined and compared to the overall proportion of early and late origins (**Figure 21**). Similar to the replication timing data in **Figure 20**, the turquoise cluster, which contains more telomeric origins, contained more inactive origins than expected. The turquoise nucleosome profile had a second NDR to the right of the ACS-containing NDR (**Figure 14**). In contrast, the pink cluster which had two adjacent

NDRs (**Figure 14**), with the second NDR to the right of the ACS, had more early origins than expected. The pink cluster was closest to transcription start sites (**Figure 17B**) indicating coding genes may influence the replication of nearby origins. The green cluster which had a distinct nucleosome occupancy pattern (**Figure 14**) contained more inactive origins than expected. Thus, different nucleosome occupancy patterns are able to explain differences in origin replication timing.

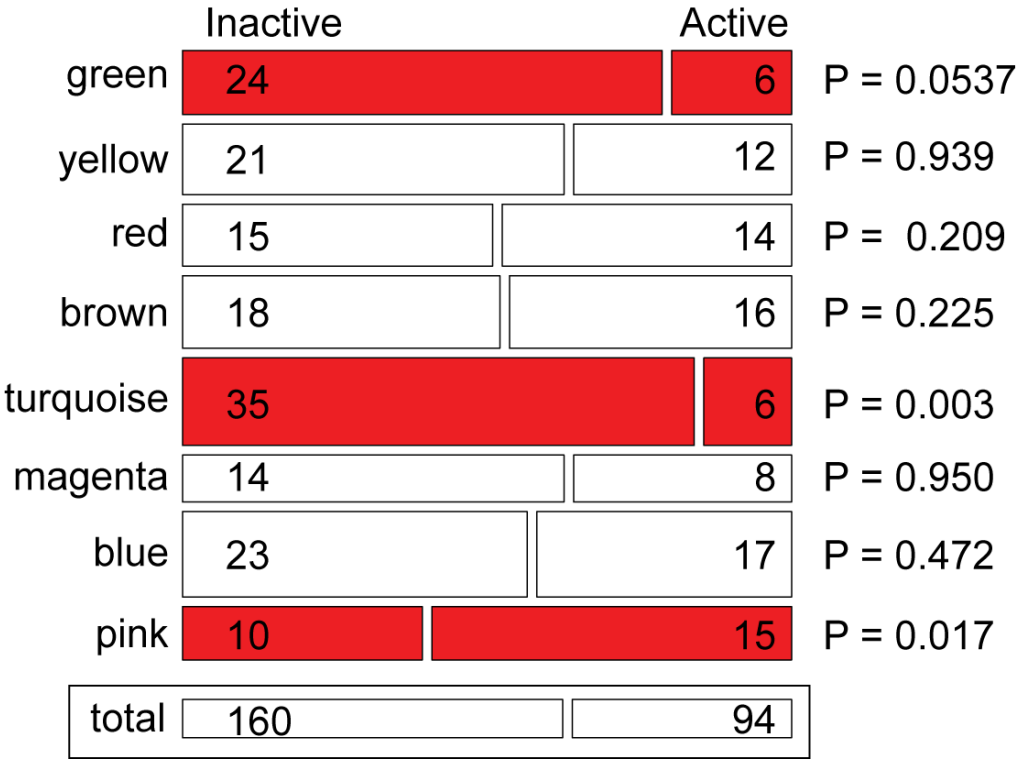

Figure 21: Origin activity in HU presented as a mosaic plot.
Origin activity in hydroxyurea data (Feng et al., 2006) was used to compare different nucleosome profile clusters. The observed proportion of early (active in HU) and late (inactive in HU) origins was compared against expected proportions using individual Chi-square tests. Significant differences are highlighted in red.

3.5 Binding of the origin recognition complex positions nucleosomes at origins

Nucleosome positioning at origins may be a consequence of ORC binding to the ACS. Using genetic perturbation of ORC it is possible to determine the role of ORC in positioning

nucleosomes adjacent to the ACS. Genetic perturbation of ORC was accomplished using an *orc2-1* allele driven by a *GAL1* promoter (Shimada et al., 2002). The *orc2-1* allele has reduced stability; it has a half-life of ~8-minutes while the wild-type protein has a half-life of ~2h (Shimada et al., 2002). Using the *GAL1* promoter, the *orc2-1* allele is tightly repressed in media containing glucose such as YPAD (Shimada and Gasser, 2007). Using *GAL:orc2-1* the Orc2 levels are depleted below the detection limit within 60 minutes (Shimada and Gasser, 2007). Depletion of Orc2 in mitosis reduces ORC function preventing DNA replication in the subsequent cell cycle (Shimada and Gasser, 2007). *GAL:orc2-1* cells accumulate in late G1 phase (**Figure 22B**) with a 1C (amount of DNA within a haploid nucleus) DNA content while wild-type cells proceed through the cell cycle and contain approximately equal proportions of cells with a 1C and 2C DNA content (**Figure 22A**).

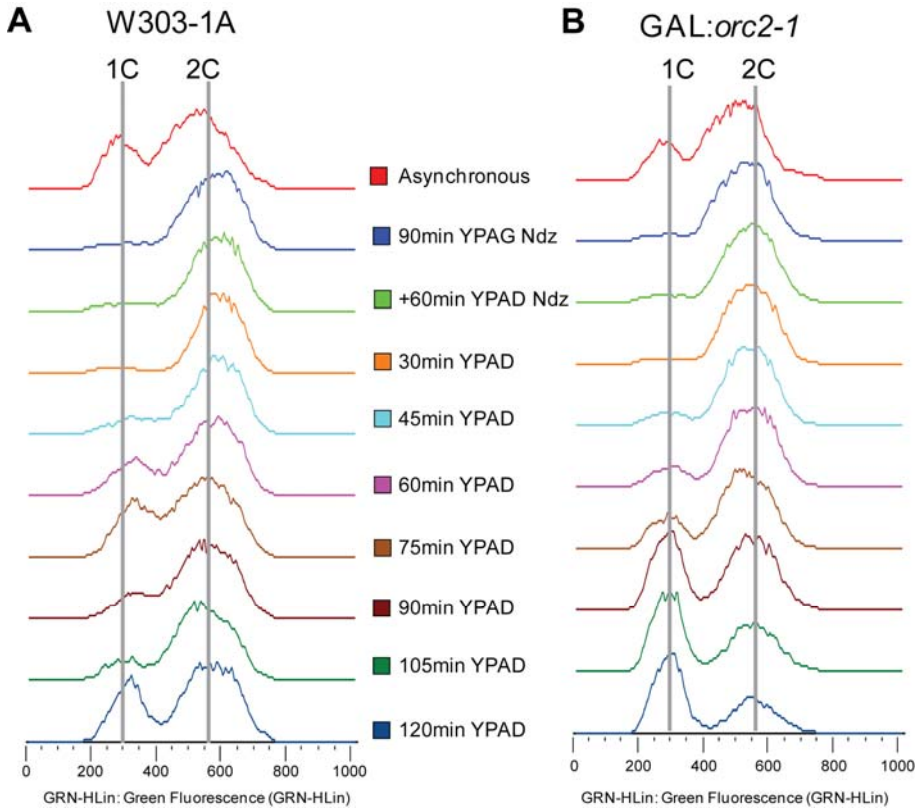

Figure 22: Depletion of Orc2 in mitosis causes a G1 arrest.
Cells were grown in a galactose-containing rich medium (YPAG) and arrested in mitosis using nocodazole. Cells were released into glucose-containing rich medium (YPAD) for 2h. The DNA content was measured using flow cytometry.

In order to determine whether nucleosome positions at origins change in response to the loss of ORC, nucleosomal DNA was isolated from *GAL:orc2-1* (2h after release from a nocodazole block into YPAD) and the congenic wild-type strain (W303-1A) and analyzed to create nucleosome maps. On average, the nucleosome depletion at origins (**Figure 23A, B**) was reduced in *GAL:orc2-1*, corresponding to a narrower NDR. The wild-type NDR was 269-bp while the *GAL:orc2-1* NDR was 217-bp (**Figure 24**). The distance between adjacent nucleosome centers were comparable between W303-1A and *GAL:orc2-1.* The nucleosome array surrounding *GAL:orc2-1* (**Figure 23B**) appears to be more delocalized, with reduced amplitude of peaks and troughs, compared to W303-1A (**Figure 23A**). The locations of nucleosomes within *GAL:orc2-1* compared to W303-1A have shifted inwards towards the ACS. This change in nucleosome positioning is highlighted by comparing the nucleosomal DNA of *GAL:orc2-1* with that of W303-1A (**Figure 23C**). These results suggest that ORC makes a strong contribution to the positioning of nucleosomes surrounding origins. In contrast to origins, the nucleosome occupancy at promoters was largely unchanged between *GAL:orc2-1* and the wild-type (**Figure 25**).

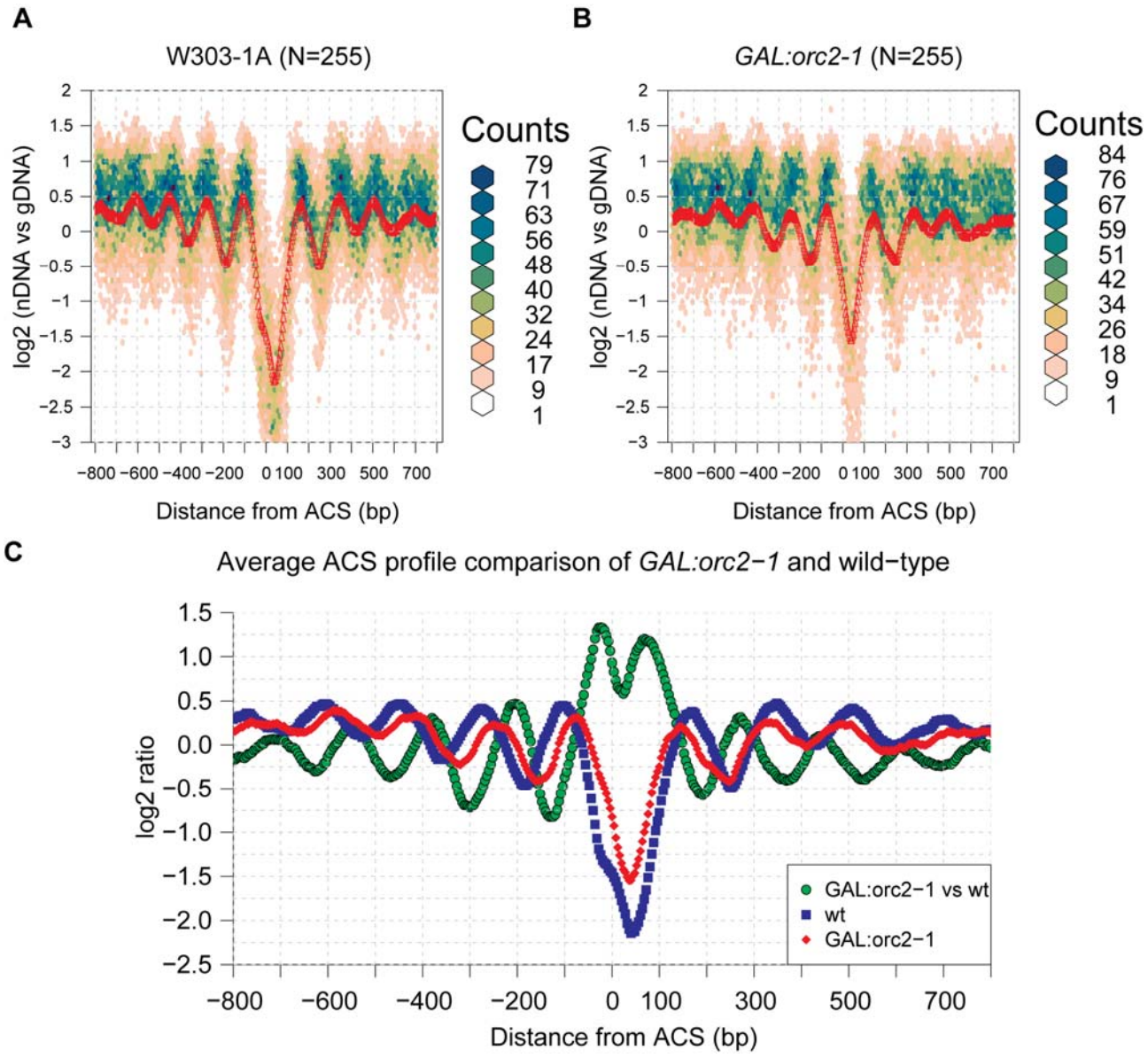

Figure 23: Nucleosome occupancy changes in *GAL:orc2-1* compared to the wild-type. The nucleosome occupancy in *GAL:orc2-1* and W303-1A are different. In W303-1A (**A**) the NDR has a larger magnitude and is wider compared to *GAL:orc2-1* (**B**). The nucleosomes have shifted inwards in *GAL:orc2-1* compared to W303-1A (**C**). The shift in nucleosome positioning is highlighted by the green nucleosome difference profile which compares nucleosomal DNA within *GAL:orc2-1* to nucleosomal DNA within W303-1A. The red and blue profiles compare ACS-centered nucleosomal DNA of *GAL:orc2-1* and W303-1A against W303-1A genomic DNA providing an indication of nucleosome positions.

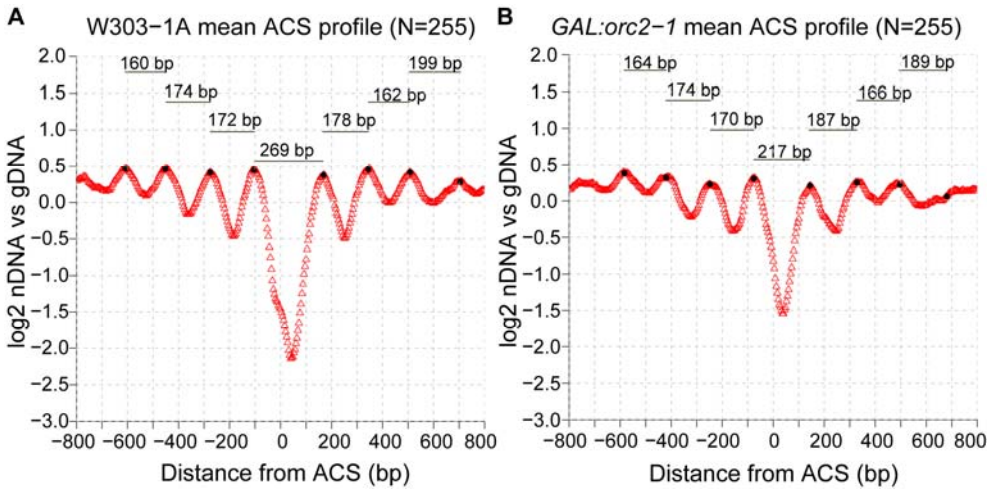

Figure 24: Comparison of NDR size between *GAL:orc2-1* and the wild-type.
The size of the nucleosome-depleted region (NDR) is reduced in *GAL:orc2-1* compared to W303-1A. The distance between nucleosome centers is similar between *GAL:orc2-1* and W303-1A.

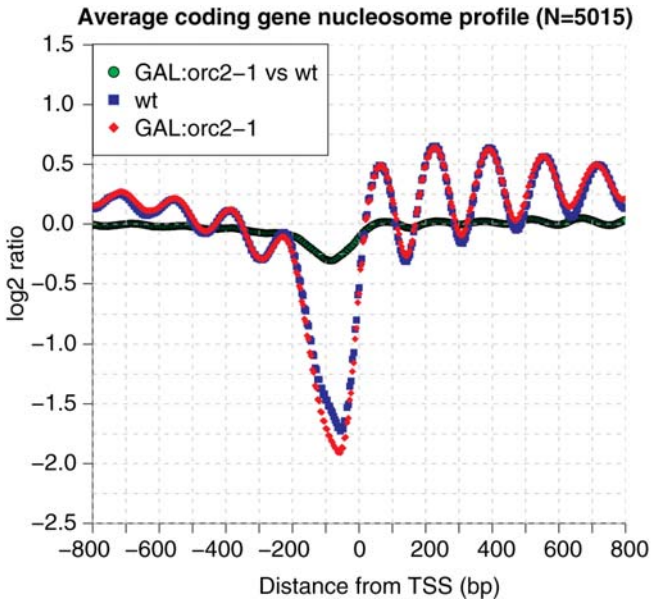

Figure 25: Average TSS-centered nucleosome occupancy of *GAL:orc2-1* and the wild-type.
Nucleosome occupancy at promoters centered by their transcription start site (TSS) is largely unchanged between *GAL:orc2-1* and the wild-type.

Despite Orc2 becoming fully depleted within 60 minutes of transferring *GAL:orc2-1* to media containing glucose, residual Orc2 may remain protected within the pre-RC (Shimada and Gasser, 2007). Using clustering analysis it was possible to determine which origins were most affected by ORC depletion. Clustering revealed two main groups: one group in which there were changes in nucleosome occupancy at the ACS and another group with minor changes in nucleosome occupancy at the ACS (**Figure 26**). In cluster#2 (**Figure 26**) nucleosomes to the left of the ACS were shifted inwards towards the ACS. Nucleosomes to the right of the ACS-containing NDR appear to become delocalized; the peak-to-trough amplitude is reduced in the mutant compared to the wild-type. Whether these 2 groups possess different amounts of residual Orc2 remains to be determined by performing a ChIP-chip experiment with *GAL:orc2-1*.

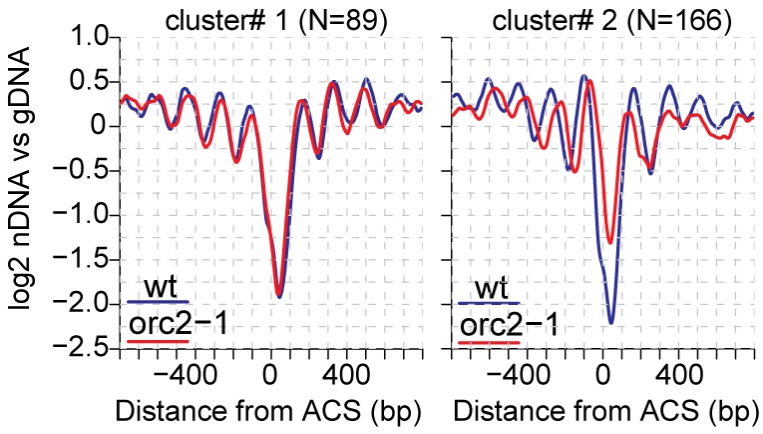

Figure 26: Orc2 depletion has a significant influence on origin nucleosome architecture. The difference between *GAL:orc2-1* and wild-type nucleosomal DNA was clustered into 2 groups using k-means clustering. The average nucleosome occupancy for origins in cluster#1 are similar between the wild-type and mutant. Cluster#2 origins are shifted inward towards the ACS and the magnitude of the NDR is reduced in the mutant compared to the wild-type.

Using the wild-type clusters of nucleosome occupancy surrounding the ACS in **Figure 13** it was possible to identify which groups of origins experienced changes in nucleosome occupancy following Orc2 depletion (**Figure 27**). In **Figure 27A** the differences in nucleosome occupancy between *GAL:orc2-1* and the wild-type are shown. The turquoise cluster which was found to contain subtelomeric origins experienced a substantial increase in nucleosome occupancy within the ACS-containing NDR following Orc2 depletion. Generally, nucleosomes shift inward towards the ACS-containing NDR and the size of the ACS-containing NDR is reduced when comparing *GAL:orc2-1* nucleosome occupancy (**Figure 27B**) to wild-type nucleosome occupancy (**Figure 27C**). The differences between *GAL:orc2-1* and the wild-type nucleosome architecture is easier to visualize using a subcluster average view (**Figure 28**). The green cluster lacks a large ACS-containing NDR in both *GAL:orc2-1* and the wild-type. The size of the ACS-containing NDR is reduced in *GAL:orc2-1* compared to wild-type. In the yellow and brown clusters the nucleosomes to the left of the ACS are shifted inward towards the ACS and the phasing of nucleosomes to the right of the ACS is reduced. In the red cluster, nucleosomes to the left of the ACS are shifted inward towards the ACS but the nucleosomes to the right of the ACS are unchanged when comparing the mutant to the wild-type. The turquoise and magenta clusters (**Figure 28)** have the largest change in nucleosome occupancy: the magnitude of the depletion at the NDR is reduced and positioned nucleosomes to the left and right of the ACS move inward towards the ACS. In the blue cluster the magnitude of the ACS-containing NDR is reduced and nucleosomes on either side of the ACS are shifted inward towards the ACS when comparing the mutant against the wild-type. Finally, the pink cluster which contained a unique dual NDR profile had a significant reduction in the magnitude of the ACS-containing NDR and nucleosomes to the right of the ACS are shifted inward towards the ACS. The magnitude of the NDR to the left of the ACS was slightly increased when comparing the mutant to the wild-type

and the positioning of the nucleosome between the two NDRs was unchanged. In general, the subcluster average view in **Figure 28** reveals that nucleosome positioning changes following ORC depletion involve nucleosomes shifting positions or becoming more delocalized. These changes indicate that nucleosomes were no longer positioned by ORC and were able to move inward towards the ACS.

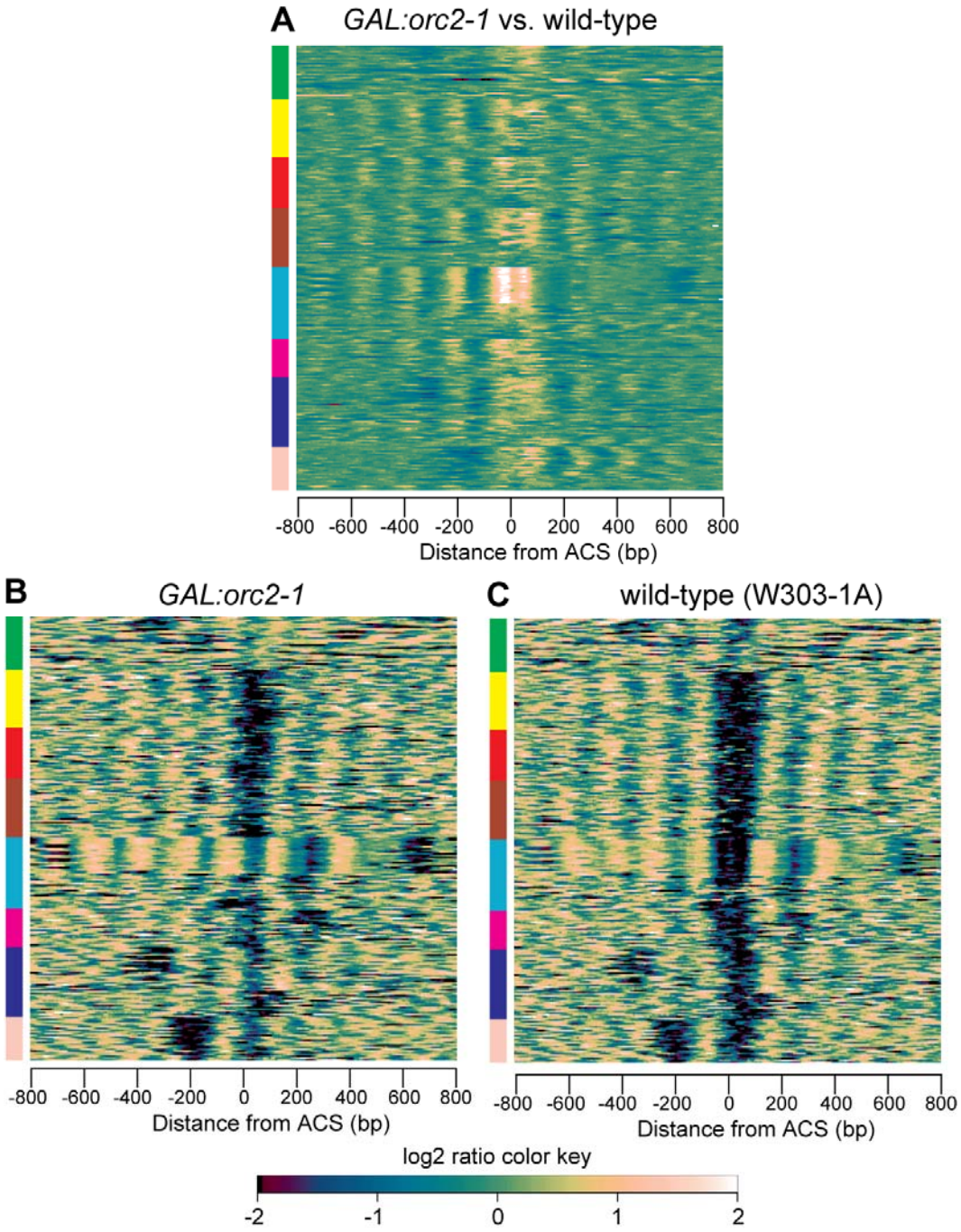

Figure 27: Heatmap highlighting differences in nucleosome occupancy between *GAL:orc2-1* and the wild-type.
Nucleosome occupancy differences between *GAL:orc2-1*and the wild-type (W303-1A) are grouped according to the clusters shown in **Figure 13**. The origins within each group are sorted by their similarity to the average difference in nucleosome occupancy between *GAL:orc2-1* nucleosomal DNA and wild-type nucleosomal DNA (**A**). *GAL:orc2-1* (**B**) and wild-type (**C**) nucleosome occupancy was compared against wild-type genomic DNA.

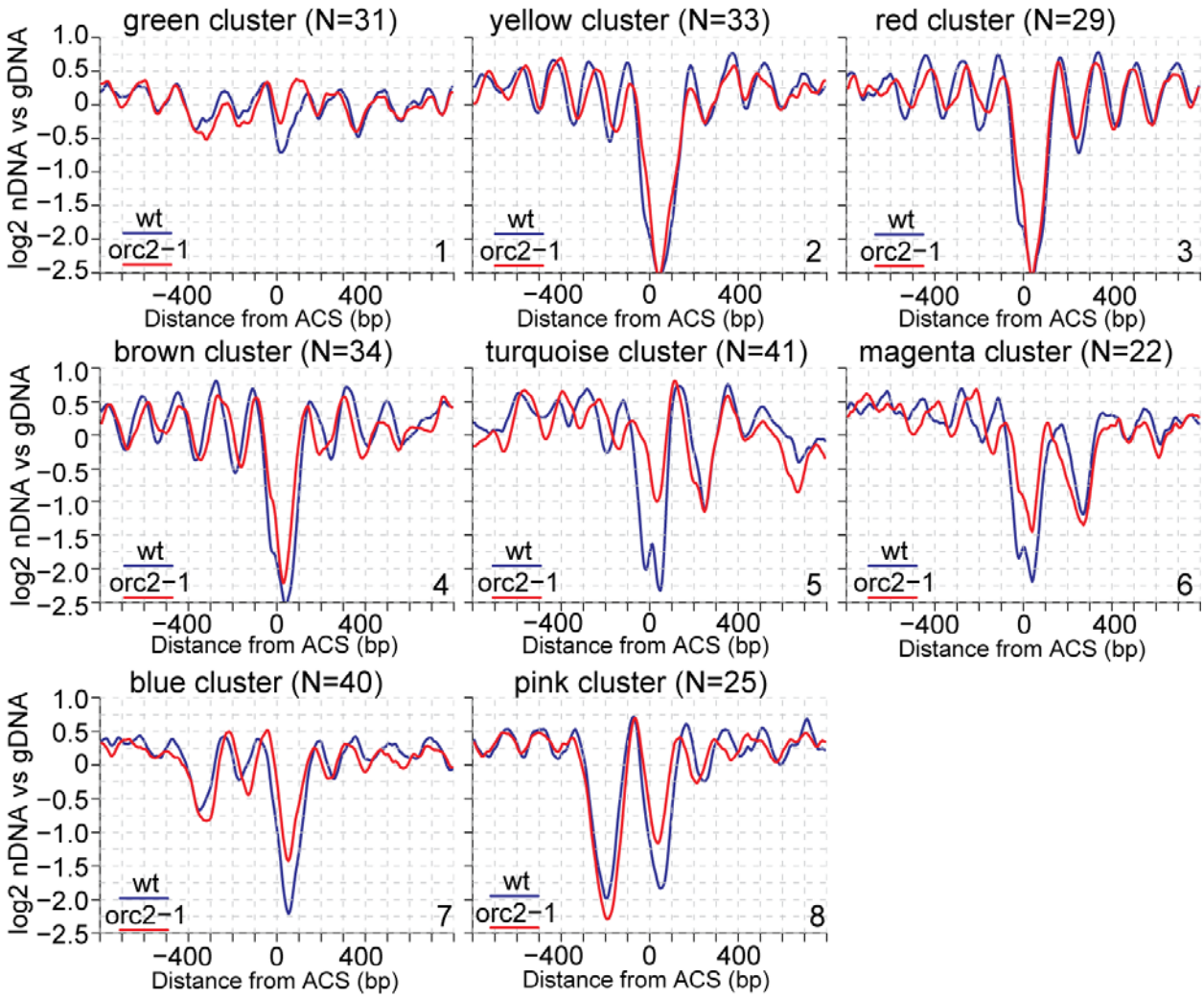

Figure 28: Subclusters highlighting differences between *GAL:orc2-1* and the wild-type nucleosome profiles.
Each panel presents a comparison between the nucleosome occupancy of *GAL:orc2-1* and the wild-type for each subcluster shown **Figure 27.** Each plot was smoothed in a 5-probe (20-bp) window. In general, nucleosome occupancy changes occur at the ACS-containing NDR or the positioning and/or phasing of adjacent nucleosomes. See main text for details.

3.6 The ACS remains nucleosome-free when chromatin is assembled *in vitro*

The size of the NDR at the ACS was reduced, but not eliminated, upon Orc2 depletion. One explanation for the modest effect is that the NDR containing the ACS may contain sequence encoded nucleosome exclusion signals (Field et al., 2008). Alternatively, incomplete inactivation of ORC may prevent the ACS from becoming fully nucleosome occupied. Using *in vitro*

nucleosome maps (Kaplan et al., 2009) it is possible to distinguish between these two alternatives. *In vitro* nucleosome maps indicate the intrinsic sequence preferences of nucleosomes without the added complexity of other non-histone DNA binding proteins. The average ACS-centered profile of 198 ARSs (**Figure 29**) indicated that the region surrounding the ACS is a sequence encoded NDR; a region of ~400-bp. To the left and right of the ACS there are no positioned nucleosomes, indicating that nucleosomes surrounding the ACS are not sequence encoded. This is reminiscent of the promoter architecture in these same samples. The ~400-bp NDR is larger than observed *in vivo,* indicating ORC and other non-histone DNA-binding proteins contribute to the generation of an array of phased nucleosomes surrounding the ACS.

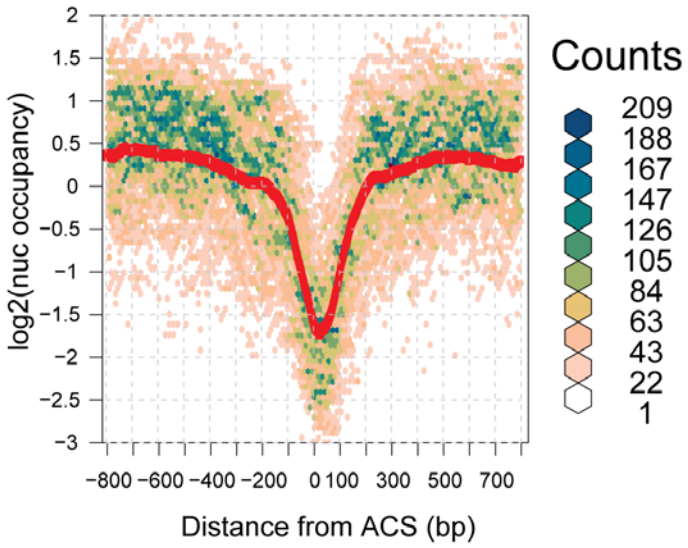

Figure 29: *In vitro* ACS-centered nucleosome profile.
The average ACS-centered nucleosome profile was extracted from 198 origins. The origins were obtained from Kaplan et al. as described in Materials and Methods. There is a ~400-bp NDR; a region with a nucleosome occupancy less than 0. There are no positioned nucleosomes to the left and right of the ACS.

Chapter 4
Discussion and Future Directions

My analysis of ACS-centered nucleosomes is distinct from previous genome-wide investigations of nucleosome occupancy at origins. Using nucleosome maps aligned by a set of 255 ORC-binding sites (ACSs) allowed the detection of the ACS-containing NDR and flanking nucleosomes previously reported. In contrast to previous reports, my analysis of nucleosome occupancy for origins centered on the ACS revealed that ACSs are generally located within a nucleosome-depleted region (NDR) surrounded on either side by well-positioned nucleosomes. On average, the nucleosome organization at origins is symmetric with 3 to 4 nucleosomes on either side of the ACS-containing NDR. This organization is distinct from nucleosome organization at promoters in which an array of positioned nucleosomes extends in the direction of the open reading frame.

Nucleosome organization at promoters correlates with DNA sequence features. Using average GC-content surrounding ACS-centered origins I was able to show that the ACS lies within an AT-rich region. The region with the lowest GC-content encompassed the ACS-containing nucleosome-depleted region. Investigating 103 DNA dinucleotide properties I determined that most DNA sequence features can explain the ACS-containing NDR but cannot explain the locations of positioned nucleosomes.

Differences in origin structure were highlighted by the identification of 8 distinct nucleosome profiles using hierarchical clustering. Distinct nucleosome occupancy patterns included: origins without an extended ACS-containing NDR, origins with a second NDR to the right of the ACS-containing NDR and a set of origins with a second NDR to the left of the ACS-containing NDR. The 8 classes of origins were used to compare origin properties: motif-content, genomic-context,

and origin activity. Comparing motif-content between the 8 origin classes revealed there were only minor changes in the information content of the ACS sequence and the B1-element between clusters. One class of origins, which had a NDR to the right of the ACS-containing NDR, was found to contain more information content in the region between the ACS and the B1 element. This indicated that origins within this cluster (turquoise) contained more repetitive DNA. By performing origin location analysis I determined that this cluster contained subtelomeric origins which tend to have repetitive DNA. The genomic-context comparison of different origin classes provided further insight into other nucleosome profiles, e.g., origins which contained a NDR to the left of the ACS-containing NDR (pink cluster) were the closest to transcription start sites. I also determined that origins which lack an extensive ACS-containing NDR had the closest proximity to adjacent origins. This may indicate that these origins are less efficient; the unlicensed form of ORC may predominate at these origins. My investigation into the motif-content and genomic-context of origins provides a framework to explain differences in origin activity based on their nucleosome profile.

The main goal of analyzing nucleosome profiles was to determine whether or not differences in origin activity are explained by differences in chromatin structure. Using replication timing data I found that the replication time of origins containing a NDR to the right of the ACS-containing NDR tended to have a later replication time. The late replication time of these origins correlated with the presence of subtelomeric origins. Unfortunately, differences in replication time do not distinguish between origins with a NDR to the left of the ACS and origins with a profile matching the average ACS profile. Using a different origin activity metric, origin activity in HU, I was able to show that origins containing a NDR to the right of the ACS had more late origins than expected while origins with a NDR to the left of the ACS contained more early origins than expected. Origins which lacked an extensive ACS-containing NDR had more late origins than

expected providing support for the idea that most of these origins are less efficient than other origins within this dataset. By analyzing origin activity of different nucleosome classes I was able to show that origins with distinct nucleosome architectures correspond to origins with distinct biological activities.

Single gene studies have shown that Abf1 has a role in establishing chromatin structure at origins. It is possible that differences in nucleosome architecture, specifically, the second NDR to the left or right of the ACS are a result of Abf1 binding sites. I found the locations of Abf1 binding sites within the 1600-bp region surrounding the ACS. Most Abf1 binding sites were located ~230-bp to the right of the ACS and were found within the subtelomeric turquoise cluster which had a second NDR to the right of the ACS. The factor(s) responsible for the profiles containing a second NDR to the left of the ACS remain unknown. Given the proximity of this cluster to promoters which usually contain an Abf1 binding site it was surprising that Abf1 binding sites were not identified to left of the ACS-containing NDR.

The statistical positioning of nucleosomes explains most of the nucleosome occupancy at origins. The barrier against which nucleosomes are packaged is the ACS-containing NDR in which ORC binds the ACS. The precise phasing of nucleosomes adjacent to the ACS-containing NDR is heavily influenced by ORC. Distal to the first nucleosome on either side of this barrier nucleosomes occupancy is more diffuse. Genetically perturbing ORC (which has a role in positioning nucleosome surrounding the ACS) resulted in a shift in nucleosome positions. I determined the locations of nucleosomes after ORC depletion and compared these locations to wild-type nucleosome locations. I determined that the size of the ACS-containing NDR was reduced following ORC depletion. The changes in nucleosome occupancy were limited to a subset of origins (N=166) indicating that residual Orc2 may remain at the set of origins not

experiencing changes in nucleosome occupancy (N=89). Using the 8 nucleosome classes which describe distinct nucleosome architectures I determined that unaffected origins were distributed throughout the 8 nucleosome classes. There were three types of nucleosome occupancy changes when comparing mutant and wild-type nucleosome positions: (1) a shift in nucleosome positions on the left-side of the ACS; (2) a shift in nucleosome positions on the right-side of the ACS; and (3) increased nucleosome occupancy at the ACS-containing NDR. My observation that nucleosomes shifted inward towards the ACS and became more delocalized indicates ORC plays a strong role in positioning nucleosomes adjacent to the ACS.

ORC depletion did not result in the loss of the ACS-containing NDR. Using a dataset describing the locations of nucleosomes loaded onto purified yeast genomic DNA (*in vitro* nucleosome locations) I determined that the region surrounding the ACS was a sequence-encoded NDR. The sequence-encoded NDR is larger than the NDR observed *in vivo* indicating that ORC and other DNA-binding proteins generate the *in vivo* nucleosome occupancy pattern. The size of this NDR is reduced in the absence of ORC because ORC keeps nucleosomes at precise positions surrounding the ACS. In the absence of ORC the positioning of these nucleosomes is no longer constrained and they move (as a result of nucleosome sliding and/or chromatin remodelling) as close as possible to the remaining barrier: a sequence of nucleosome excluding bases. The NDR creates an environment in which ORC and other pre-RC components can easily bind to the underlying DNA. Once bound to the pre-RC chromatin remodellers may be recruited by ORC (such as Rpd3) leading to nucleosomes moving towards the NDR. The nucleosomes adjacent to ORC may play a role in recruiting MCM proteins to the pre-RC (Lipford and Bell, 2001). Thus, larger *in vivo* NDRs may correspond to less efficient origins. The novel findings presented in this study include all of the information derived from the average view of replication origins, the discovery of a previously unappreciated diversity of nucleosome structure at origins, a

statistically robust clustering analysis that provides biological insight into the relationship between origin structure and function, and genome-wide analysis of the effect of ORC depletion on nucleosome positioning.

Future work will involve investigating mutants which may have a role in positioning nucleosomes at origins. Mcm10 has a role in the initiation of DNA replication and the progression of replication forks, as a *mcm10-1* mutant pauses replication forks adjacent to origins of replication (Kawasaki et al., 2000). Given these two roles Mcm10 may function at the transition from initiation to elongation (Bell and Dutta, 2002). Obtaining nucleosomes from a *mcm10-1* mutant arrested with α-factor at the non-permissive temperature (37°C) and then released could reveal changes in nucleosome occupancy at origins associated with the disassembly of the pre-replicative complex (Kawasaki et al., 2000).

Mcm1 is a transcription factor which regulates the expression of some DNA replication genes (Tye, 1999). Mcm1 may influence the chromatin structure of replication origins by binding to sites which overlap origin B3 elements (in *ARS1* and *ARS121*) (Chang et al., 2003). The B3 element is usually considered to be an Abf1 binding site, but Abf1 binding to the B3 element of *ARS1* has been shown *in vitro* but not *in vivo* and an *abf1-1* mutant does not effect *ARS1* firing (Chang et al., 2003). Therefore, obtaining nucleosomes from *mcm1-1* at the non-permissive temperature, and observing the nucleosome structure at origins may reveal the cause of origins containing two nucleosome-depleted regions, these origins may contain Mcm1 binding sites.

Additional work with mutants which influence late origin firing may reveal nucleosome occupancy patterns which explain why some origins are early while others are late. Rpd3, a histone deacetylase, delays the replication of many late-origins (Aparicio et al., 2004). Obtaining Δ*rpd3* nucleosomes, in which late origins are activated early, and searching for changes in

nucleosome occupancy at origins in comparison to the wild-type may reveal the nucleosome signature of late origins and the nucleosome positioning changes needed for these origins to become early. In addition, differences between early and late origins may be revealed by obtaining Δ*clb5* nucleosomes. A *CLB5* deletion strain has a longer S-phase which is associated with significant delays in origin firing (McCune et al., 2008). Origins which fire in late S-phase have the largest delay in replication timing (McCune et al., 2008). This phenotype may enhance the differences in nuleosome structure between early and late origins revealing a unique signature of nucleosome occupancy at late origins. Finally, obtaining nucleosomes from cells lacking Mec1 and Rad53, kinases involved in the intra-S checkpoint which senses DNA damage and incomplete DNA replication, may reveal differences between the nucleosome signatures of early and late origins (Tye, 1999). Late origins replicate early in the absence of Mec1 and Rad53 (Tye, 1999). Obtaining nucleosomes from each of these mutants should definitively resolve whether or not early and late origins have distinct nucleosome architectures.

In order to further refine our knowledge of nucleosome structure at origins in *S. cerevisiae* it is necessary to identify and confirm the ORC-binding site (ACS) for each of the ~732 origins (Nieduszynski et al., 2007). This involves performing many site-directed mutagenesis experiments. A quicker method to identify ORC binding sites and to refine the area over which the ACS may be localized is to identify regions in the genome which contain ORC-positioned nucleosomes. Such sites can be identified based on the architecture of ORC-positioned nucleosomes: ~100-bp nucleosome-depleted region bordered by 2 well positioned nucleosomes. A major challenge will be to extend nucleosome positioning analysis in yeast to other eukaryotes. As a starting point it would be interesting to determine if other *sensu stricto Saccharomyces* species contain similar nucleosome organization at their origins of replication. The importance of DNA sequence in specifying DNA replication origins is reduced in higher

eukaryotes; nucleosome positioning adjacent to ORC binding sites will be a particularly useful analysis to determine the locations and differences among origins.

References

Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature *446*, 572-576.

Ambrose, C., Lowman, H., Rajadhyaksha, A., Blasquez, V., and Bina, M. (1990). Location of nucleosomes in simian virus 40 chromatin. J Mol Biol *214*, 875-884.

Anderson, J.D., and Widom, J. (2000). Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. J Mol Biol *296*, 979-987.

Aparicio, J.G., Viggiani, C.J., Gibson, D.G., and Aparicio, O.M. (2004). The Rpd3-Sin3 histone deacetylase regulates replication timing and enables intra-S origin control in Saccharomyces cerevisiae. Mol Cell Biol *24*, 4769-4780.

Aparicio, O.M., Stout, A.M., and Bell, S.P. (1999). Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. Proc Natl Acad Sci U S A *96*, 9130-9135.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L.*, et al.* (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol Cell *32*, 878-887.

Becker, P.B. (2002). Nucleosome sliding: facts and fiction. EMBO J *21*, 4749-4753.

Bell, S.P. (1995). Eukaryotic replicators and associated protein complexes. Curr Opin Genet Dev *5*, 162-167.

Bell, S.P., and Dutta, A. (2002). DNA replication in eukaryotic cells. Annu Rev Biochem *71*, 333-374.

Bell, S.P., and Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128-134.

Blow, J.J., and Dutta, A. (2005). Preventing re-replication of chromosomal DNA. Nat Rev Mol Cell Biol *6*, 476-486.

Breier, A.M., Chatterji, S., and Cozzarelli, N.R. (2004). Prediction of Saccharomyces cerevisiae replication origins. Genome Biol *5*, R22.

Carr, D., Lewin-Koh, N., and Maechler, M. (2009). hexbin: Hexagonal Binning Routines.

Chang, V.K., Fitch, M.J., Donato, J.J., Christensen, T.W., Merchant, A.M., and Tye, B.K. (2003). Mcm1 binds replication origins. J Biol Chem *278*, 6093-6100.

Charif, D., and Lobry, J.R. (2007). SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In Structural approaches to sequence evolution: Molecules, networks, populations (New York, Springer Verlag), pp. 207-232.

Chesnokov, I.N. (2007). Multiple functions of the origin recognition complex. Int Rev Cytol *256*, 69-109.

Chou, T. (2007). Peeling and sliding in nucleosome repositioning. Phys Rev Lett *99*, 058105.

Crampton, A., Chang, F., Pappas, D.L., Jr., Frisch, R.L., and Weinreich, M. (2008). An ARS element inhibits DNA replication through a SIR2-dependent mechanism. Mol Cell *30*, 156-166.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Czajkowsky, D.M., Liu, J., Hamlin, J.L., and Shao, Z. (2008). DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. J Mol Biol *375*, 12-19.

Dahmann, C., Diffley, J.F., and Nasmyth, K.A. (1995). S-phase-promoting cyclin-dependent kinases prevent re-replication by inhibiting the transition of replication origins to a pre-replicative state. Curr Biol *5*, 1257-1269.

Davierwala, A.P., Haynes, J., Li, Z., Brost, R.L., Robinson, M.D., Yu, L., Mnaimneh, S., Ding, H., Zhu, H., Chen, Y.*, et al.* (2005). The synthetic genetic interaction spectrum of essential genes. Nat Genet *37*, 1147-1152.

Diller, J.D., and Raghuraman, M.K. (1994). Eukaryotic replication origins: control in space and time. Trends Biochem Sci *19*, 320-325.

Eddelbuettel, D. (2009). random: True random numbers using random.org.

Elsasser, S., Chi, Y., Yang, P., and Campbell, J.L. (1999). Phosphorylation controls timing of Cdc6p destruction: A biochemical analysis. Mol Biol Cell *10*, 3263-3277.

Ercan, S., and Lieb, J.D. (2006). New evidence that DNA encodes its packaging. Nat Genet *38*, 1104-1105.

Fangman, W.L., Hice, R.H., and Chlebowicz-Sledziewska, E. (1983). ARS replication during the yeast S phase. Cell *32*, 831-838.

Feng, W., Collingwood, D., Boeck, M.E., Fox, L.A., Alvino, G.M., Fangman, W.L., Raghuraman, M.K., and Brewer, B.J. (2006). Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. Nat Cell Biol *8*, 148-155.

Field, Y., Fondufe-Mittendorf, Y., Moore, I.K., Mieczkowski, P., Kaplan, N., Lubling, Y., Lieb, J.D., Widom, J., and Segal, E. (2009). Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. Nat Genet *41*, 438-445.

Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I.K., Sharon, E., Lubling, Y., Widom, J., and Segal, E. (2008). Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. PLoS Comput Biol *4*, e1000216.

FitzGerald, P.C., and Simpson, R.T. (1985). Effects of sequence alterations in a DNA segment containing the 5 S RNA gene from Lytechinus variegatus on positioning of a nucleosome core particle in vitro. J Biol Chem *260*, 15318-15324.

Friedel, M., Nikolajewa, S., Suhnel, J., and Wilhelm, T. (2009). DiProDB: a database for dinucleotide properties. Nucleic Acids Res *37*, D37-40.

Friedman, K.L., Diller, J.D., Ferguson, B.M., Nyland, S.V., Brewer, B.J., and Fangman, W.L. (1996). Multiple determinants controlling activation of yeast replication origins late in S phase. Genes Dev *10*, 1595-1607.

Hartwell, L. (1992). Defects in a cell cycle checkpoint may be responsible for the genomic instability of cancer cells. Cell *71*, 543-546.

Hartwell, L.H., Culotti, J., Pringle, J.R., and Reid, B.J. (1974). Genetic control of the cell division cycle in yeast. Science *183*, 46-51.

Hartwell, L.H., Culotti, J., and Reid, B. (1970). Genetic control of the cell-division cycle in yeast. I. Detection of mutants. Proc Natl Acad Sci U S A *66*, 352-359.

Hayes, J.J., and Wolffe, A.P. (1992). The interaction of transcription factors with nucleosomal DNA. Bioessays *14*, 597-603.

Hirschman, J.E., Balakrishnan, R., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hong, E.L., Livstone, M.S., Nash, R.*, et al.* (2006). Genome Snapshot: a new resource at the Saccharomyces Genome Database (SGD) presenting an overview of the Saccharomyces cerevisiae genome. Nucleic Acids Res *34*, D442-445.

Huberman, J.A., and Riggs, A.D. (1968). On the mechanism of DNA replication in mammalian chromosomes. J Mol Biol *32*, 327-341.

Ioshikhes, I.P., Albert, I., Zanton, S.J., and Pugh, B.F. (2006). Nucleosome positions predicted through comparative genomics. Nat Genet *38*, 1210-1215.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet *10*, 161-172.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J.*, et al.* (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.

Kawasaki, Y., Hiraga, S., and Sugino, A. (2000). Interactions between Mcm10p and other replication factors are required for proper initiation and elongation of chromosomal DNA replication in Saccharomyces cerevisiae. Genes Cells *5*, 975-989.

Keich, U., Gao, H., Garretson, J.S., Bhaskar, A., Liachko, I., Donato, J., and Tye, B.K. (2008). Computational detection of significant variation in binding affinity across two sets of sequences with application to the analysis of replication origins in yeast. BMC Bioinformatics *9*, 372.

Knott, S.R.V., Viggiani, C.J., TavarÃ©, S., and Aparicio, O.M. (2009). Genome-wide replication profiles indicate an expansive role for Rpd3L in regulating replication initiation timing or efficiency, and reveal genomic loci of Rpd3 function in Saccharomyces cerevisiae. Genes & Development *23*, 1077-1090.

Kornberg, R. (1981). The location of nucleosomes in chromatin: specific or statistical. Nature *292*, 579-580.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.

Kornberg, R.D., and Lorch, Y. (1992). Chromatin structure and transcription. Annu Rev Cell Biol *8*, 563-587.

Kornberg, R.D., and Stryer, L. (1988). Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. Nucleic Acids Res *16*, 6677-6690.

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics *24*, 719-720.

Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet *36*, 900-905.

Lee, D.G., and Bell, S.P. (1997). Architecture of the yeast origin recognition complex bound to origins of DNA replication. Mol Cell Biol *17*, 7159-7168.

Lee, D.Y., Hayes, J.J., Pruss, D., and Wolffe, A.P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell *72*, 73-84.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet *39*, 1235-1244.

Lipford, J.R., and Bell, S.P. (2001). Nucleosomes positioned by ORC facilitate the initiation of DNA replication. Mol Cell *7*, 21-30.

Louis, E.J. (1995). The chromosome ends of Saccharomyces cerevisiae. Yeast *11*, 1553-1573.

Lucas, A. (2009). amap: Another Multidimensional Analysis Package.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

MacAlpine, D.M., and Bell, S.P. (2005). A genomic view of eukaryotic DNA replication. Chromosome Res *13*, 309-326.

MacIsaac, K.D., and Fraenkel, E. (2006). Practical strategies for discovering regulatory DNA sequence motifs. PLoS Comput Biol *2*, e36.

Maechler, M., Rousseeuw, P., Struyf, A., and Hubert, M. (2005). Cluster Analysis Basics and Extensions.

Marahrens, Y., and Stillman, B. (1992). A yeast chromosomal origin of DNA replication defined by multiple functional elements. Science *255*, 817-823.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res *18*, 1073-1083.

McCarroll, R.M., and Fangman, W.L. (1988). Time of replication of yeast centromeres and telomeres. Cell *54*, 505-513.

McCune, H.J., Danielson, L.S., Alvino, G.M., Collingwood, D., Delrow, J.J., Fangman, W.L., Brewer, B.J., and Raghuraman, M.K. (2008). The temporal program of chromosome replication: genomewide replication in clb5{Delta} Saccharomyces cerevisiae. Genetics *180*, 1833-1847.

Meyer, D., Zeileis, A., and Hornik, K. (2009). vcd: Visualizing Categorical Data. R package version 1.2-4.

Mimura, S., and Takisawa, H. (1998). Xenopus Cdc45-dependent loading of DNA polymerase alpha onto chromatin under the control of S-phase Cdk. EMBO J *17*, 5699-5707.

Moldovan, G.L., Pfander, B., and Jentsch, S. (2007). PCNA, the maestro of the replication fork. Cell *129*, 665-679.

Nguyen, V.Q., Co, C., and Li, J.J. (2001). Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. Nature *411*, 1068-1073.

Nieduszynski, C.A., Blow, J.J., and Donaldson, A.D. (2005). The requirement of yeast replication origins for pre-replication complex proteins is modulated by transcription. Nucleic Acids Res *33*, 2410-2420.

Nieduszynski, C.A., Hiraga, S., Ak, P., Benham, C.J., and Donaldson, A.D. (2007). OriDB: a DNA replication origin database. Nucleic Acids Res *35*, D40-46.

Nieduszynski, C.A., Knox, Y., and Donaldson, A.D. (2006). Genome-wide identification of replication origins in yeast by comparative genomics. Genes Dev *20*, 1874-1879.

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

Pazin, M.J., Bhargava, P., Geiduschek, E.P., and Kadonaga, J.T. (1997). Nucleosome mobility and the maintenance of nucleosome positioning. Science *276*, 809-812.

Peckham, H.E., Thurman, R.E., Fu, Y., Stamatoyannopoulos, J.A., Noble, W.S., Struhl, K., and Weng, Z. (2007). Nucleosome positioning signals in genomic DNA. Genome Res *17*, 1170-1177.

Piatti, S., Bohm, T., Cocker, J.H., Diffley, J.F., and Nasmyth, K. (1996). Activation of S-phase-promoting CDKs in late G1 defines a "point of no return" after which Cdc6 synthesis cannot promote DNA replication in yeast. Genes Dev *10*, 1516-1531.

R Development Core Team (2009). R: A Language and Environment for Statistical Computing (Vienna, Austria).

Raghuraman, M.K., Winzeler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., and Fangman, W.L. (2001). Replication dynamics of the yeast genome. Science *294*, 115-121.

Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J., and Madhani, H.D. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. Cell *123*, 233-248.

Rando, O.J. (2007). Chromatin structure in the genomics era. Trends Genet *23*, 67-73.

Remus, D., and Diffley, J.F. (2009). Eukaryotic DNA replication control: Lock and load, then fire. Curr Opin Cell Biol.

Rowley, A., Dowell, S.J., and Diffley, J.F. (1994). Recent developments in the initiation of chromosomal DNA replication: a complex picture emerges. Biochim Biophys Acta *1217*, 239-256.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772-778.

Segal, E., and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr Opin Struct Biol *19*, 65-71.

Shimada, K., and Gasser, S.M. (2007). The origin recognition complex functions in sister-chromatid cohesion in Saccharomyces cerevisiae. Cell *128*, 85-99.

Shimada, K., Pasero, P., and Gasser, S.M. (2002). ORC and the intra-S-phase checkpoint: a threshold regulates Rad53p activation in S phase. Genes Dev *16*, 3236-3252.

Shimizu, M., Roth, S.Y., Szent-Gyorgyi, C., and Simpson, R.T. (1991). Nucleosomes are positioned with base pair precision adjacent to the alpha 2 operator in Saccharomyces cerevisiae. EMBO J *10*, 3033-3041.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V.R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. PLoS Biol *6*, e65.

Simpson, R.T. (1986). Nucleosome positioning in vivo and in vitro. Bioessays *4*, 172-176.

Simpson, R.T. (1990). Nucleosome positioning can affect the function of a cis-acting DNA element in vivo. Nature *343*, 387-389.

Simpson, R.T. (1999). In vivo methods to analyze chromatin structure. Curr Opin Genet Dev *9*, 225-229.

Stevenson, J.B., and Gottschling, D.E. (1999). Telomeric chromatin modulates replication timing near chromosome ends. Genes Dev *13*, 146-151.

Stinchcomb, D.T., Struhl, K., and Davis, R.W. (1979). Isolation and characterisation of a yeast chromosomal replicator. Nature *282*, 39-43.

Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y., and Araki, H. (2007). CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. Nature *445*, 328-332.

Thastrom, A., Lowary, P.T., Widlund, H.R., Cao, H., Kubista, M., and Widom, J. (1999). Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. J Mol Biol *288*, 213-229.

Tye, B.K. (1999). MCM proteins in DNA replication. Annu Rev Biochem *68*, 649-686.

Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B.J., and Grunstein, M. (2002). Histone acetylation regulates the time of replication origin firing. Mol Cell *10*, 1223-1233.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S.*, et al.* (2009). gplots: Various R programming tools for plotting data.

Weber, J.M., Irlbacher, H., and Ehrenhofer-Murray, A.E. (2008). Control of replication initiation by the Sum1/Rfm1/Hst1 histone deacetylase. BMC Mol Biol *9*, 100.

Whitehouse, I., Rando, O.J., Delrow, J., and Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. Nature *450*, 1031-1035.

Widom, J. (2001). Role of DNA sequence in nucleosome stability and dynamics. Q Rev Biophys *34*, 269-324.

Woods, K.K., Maehigashi, T., Howerton, S.B., Sines, C.C., Tannenbaum, S., and Williams, L.D. (2004). High-resolution structure of an extended A-tract: [d(CGCAAATTTGCG)]2. J Am Chem Soc *126*, 15330-15331.

Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., and Aparicio, O.M. (2001). Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. Science *294*, 2357-2360.

Xu, W., Aparicio, J.G., Aparicio, O.M., and Tavare, S. (2006). Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. BMC Genomics *7*, 276.

Yabuki, N., Terashima, H., and Kitada, K. (2002). Mapping of early firing origins on a replication profile of budding yeast. Genes Cells *7*, 781-789.

Yin, S., Deng, W., Hu, L., and Kong, X. (2009). The impact of nucleosome positioning on the organization of replication origins in eukaryotes. Biochem Biophys Res Commun.

Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science *309*, 626-630.

Zegerman, P., and Diffley, J.F. (2007). Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. Nature *445*, 281-285.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nat Struct Mol Biol *16*, 847-852.