Genome-wide Footprinting Uncovers Epigenetic Regulatory Paradigms by Revealing the Chromatin Occupancy Landscape

by

Jason Alan Belsky

Graduate Program in Computational Biology and Bioinformatics
Duke University

Date: __________________________

Approved:

____________________________
David M. MacAlpine, Supervisor

____________________________
Alexander J. Hartemink

____________________________
L. Ryan Baugh

____________________________
Thomas Petes

____________________________
Timothy E. Reddy

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate Program in Computational Biology and Bioinformatics in the Graduate School of Duke University 2015
ABSTRACT

Genome-wide Footprinting Uncovers Epigenetic Regulatory Paradigms by Revealing the Chromatin Occupancy Landscape

by

Jason Alan Belsky

Graduate Program in Computational Biology and Bioinformatics
Duke University

Date: __________________________

Approved:

______________________________
David M. MacAlpine, Supervisor

______________________________
Alexander J. Hartemink

______________________________
L. Ryan Baugh

______________________________
Thomas Petes

______________________________
Timothy E. Reddy

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate Program in Computational Biology and Bioinformatics in the Graduate School of Duke University 2015
Abstract

Eukaryotic genomes have extensive flexibility and plasticity to modify transcription and replication programs, yielding a myriad of differentiated cell types and survival mechanisms to adverse environmental conditions. As these genomic processes require precise localization of DNA-binding factors, their dynamic temporal and spatial distributions provide dramatically different interpretations of a static genome sequence. DNA-binding factors must compete with nucleosomes, the basic subunit of chromatin, for access to the underlying DNA sequence. Even though the spatial preferences of these proteins are partially explained by DNA sequence alone, the complete genome occupancy profile has remained elusive, and we currently have a limited understanding of how DNA-binding protein configurations directly impact transcription and replication function.

Profiling the entire chromatin environment has typically required multiple experiments to capture both DNA-binding factors and nucleosomes. Here, we have extended the traditional micrococcal nuclease (MNase) digestion assay to simultaneously resolve both nucleosomes and smaller DNA-binding footprints in *Saccharomyces cerevisiae*. Visualization of protected DNA fragments revealed a nucleotide-resolution view of the chromatin architecture at individual genomic loci. We show that different MNase digestion times can capture nucleosomes partially unwrapped or complexed with chromatin remodelers. Stereotypical DNA-binding footprints are evident across all promoters, even at low-transcribed and silent genes. By aggregating
the chromatin profiles across transcription-factor–binding sites, we precisely resolve protein footprints, yielding in vivo insights into protein-DNA interactions. Together, our MNase method, in one experiment, provides an unprecedented assessment of the entire chromatin structure genome-wide.

We utilized this approach to interrogate how the replication program is regulated by the chromatin environment surrounding DNA replication initiation sites. Pre-replicative complex (pre-RC) formation commences with recruitment of the origin recognition complex (ORC) to specific locations in the genome, termed replication origins. Although successful pre-RC assembly primes each site for S-phase initiation by loading the Mcm2-7 helicase, replication origins have substantially different activation times and efficiencies. We posited that replication origin function is substantially impacted by the local chromatin environment. Here, we resolved a high-resolution ORC-dependent footprint at 269 replication origins genome-wide. Even though ORC in S. cerevisiae remains bound at replication origins throughout the cell cycle, we detected a subset of inefficient origins that did not yield a footprint until G1, suggesting a more transient ORC interaction prior to pre-RC assembly. Nucleosome movement accommodated the pre-RC–induced expansion of the ORC-dependent footprint in G1, leading to increased activation efficiency. Mcm2-7 loading is preferentially directed to one side of each replication origin, in close proximity to the origin-flanking nucleosome. Our data demonstrates that pre-RC components are assembled into multiple configurations in vivo.

We anticipate that extending chromatin occupancy profiling to many different cell types will reveal further insights into genome regulation.
To my parents, for their endless love and support.
Contents

Abstract iv
List of Figures xi
List of Abbreviations xiii
Acknowledgements xv

1 Introduction 1

1.1 Chromatin 3
   1.1.1 Nucleosome structure 3
   1.1.2 Nucleosome positioning 5
   1.1.3 Nucleosome positioning at the TSS 7
   1.1.4 Dynamic chromatin structure affects transcriptional regulation 10

1.2 Genomic approaches for studying chromatin 13
   1.2.1 Chromatin immunoprecipitation (ChIP) 13
   1.2.2 Chromatin profiling by DNA digestion 16

1.3 DNA Replication 20
   1.3.1 Pre-replicative complex (pre-RC) formation 20
   1.3.2 Origin activation 24
   1.3.3 Regulation of pre-RC assembly and origin activation 25
   1.3.4 Replication origins 26
   1.3.5 Genome-wide mapping of replication origins 28
2 Epigenome characterization at single base-pair resolution

2.1 Introduction .................................................. 36
2.2 Results ....................................................... 37
  2.2.1 Chromatin occupancy profiling via micrococcal nuclease (MNase) digestion .......................... 37
  2.2.2 Enhanced visualization of the chromatin landscape .............. 40
  2.2.3 Chromatin footprinting at the UASg reveals a non-canonical nucleosome protection .................. 43
  2.2.4 Proximal gene promoters are consistently occupied by transcription factors .......................... 46
  2.2.5 Transcription-factor binding directs nucleosome positioning ..... 49
2.3 Discussion ................................................... 56
  2.3.1 Evaluating protein-binding dynamics by varying MNase digestion times .................................. 57
  2.3.2 Chromatin architecture at gene promoters ..................... 57
  2.3.3 Small MNase-protected fragments reveal TF-binding footprints 58
2.4 Methods ...................................................... 60
  2.4.1 Data deposition ........................................... 60
  2.4.2 Chromatin isolation ....................................... 60
  2.4.3 Illumina sequencing and analysis .................................. 60
  2.4.4 Construction of individual data signal tracks .................. 61
  2.4.5 Individual gene locus plots .................................. 61

3 Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly

3.1 Introduction ................................................... 63
3.2 Results .................................................. 67
  3.2.1 Chromatin footprinting replication origins genome-wide .... 67
  3.2.2 ORC only associates stably with a subset of functional origins in G2 ........................................ 73
  3.2.3 Nucleosome dynamics in G1 promote origin activation .... 77
  3.2.4 Origin-flanking nucleosomes restrict Mcm2-7 loading ...... 82
3.3 Discussion ................................................. 89
3.4 Methods .................................................. 95
  3.4.1 Yeast strains .......................................... 95
  3.4.2 Cell synchronization .................................... 96
  3.4.3 Chromatin preparation .................................. 96
  3.4.4 Chromatin immunoprecipitation .......................... 97
  3.4.5 Early origin activation ................................... 97
  3.4.6 Sequencing library preparation .......................... 98
  3.4.7 Sequencing read alignment to the genome .................. 98
  3.4.8 Data availability ....................................... 98
  3.4.9 Construction of individual data signal tracks .............. 98
  3.4.10 Putative origin dataset ................................ 99
  3.4.11 Origin efficiency determination ........................ 99
  3.4.12 Replication timing determination ......................... 100
  3.4.13 Individual gene locus plots ............................ 100
  3.4.14 Identifying replication origin footprints ............... 101
  3.4.15 Cell-cycle–dependent nucleosome repositioning ........... 102
  3.4.16 ORC and Mcm2-7 ChIP-seq analysis ..................... 103

4 Conclusion ................................................. 104

Bibliography .................................................. 111
x
# List of Figures

1.1 Chromatin structure at the TSS .............................................. 11
1.2 ChIP-seq schematic ................................................................. 14
1.3 DNase schematic ................................................................. 17
1.4 MNase schematic ................................................................. 19
1.5 Pre-RC formation ................................................................. 22
2.1 MNase primarily digests the linker DNA regions between nucleosomes 39
2.2 MNase fragment length profile ................................................. 40
2.3 Two-dimensional visualization schematic of paired-end MNase-digested DNA fragments ................................................. 41
2.4 Visualization of the chromatin profile captures nucleosomes and small DNA-binding proteins ................................................. 42
2.5 RSC binding is evident as a partially-unwrapped nucleosome at the UASg ................................................................. 45
2.6 Average promoter chromatin architecture shows DNA-binding proteins occupy the NFR ................................................. 47
2.7 Transcription level does not dramatically impact promoter chromatin architecture ................................................................. 48
2.8 A high-resolution footprint is revealed at Abf1-binding sites ....... 51
2.9 Abf1 occupancy positions adjacent nucleosomes ....................... 53
2.10 Two distinct chromatin configurations impact Rap1 residency .... 54
2.11 Cofactor occupancy coincides with longer Rap1 residency ........ 56
3.1 Identification of an ORC-dependent footprint by MNase mapping .. 68
3.2 Presence of ORC-dependent footprint at replication origins genome-wide 69
3.3 orc1-161 chromatin differs specifically at replication origins . . . . . 70
3.4 ORC only associates with a subset of putative origins . . . . . . . . . 71
3.5 Transient ORC association at a subset of replication origins . . . . . 72
3.6 Replication origins increase protein occupancy in G1 . . . . . . . . . 73
3.7 cdc6-1 is defective in Mcm2-7 loading . . . . . . . . . . . . . . . . . 74
3.8 Transient ORC association does not impair Mcm2-7 loading . . . . . 75
3.9 ACS sequence is not predictive of occupancy differences between the G1 & G2 Footprint and G1-Only Footprint classes . . . . . . . . . . . 76
3.10 ORC-chromatin association in G2 is a determinant of origin efficiency 77
3.11 Replication origin chromatin architecture is cell-cycle–dependent . . 78
3.12 Nucleosome movement is dependent on pre-RC formation . . . . . 79
3.13 Dynamic nucleosome position predicts ORC-dependent footprint expansion direction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80
3.14 Cell-cycle–dependent nucleosome dynamics promote origin efficiency . 81
3.15 Mcm2-7 loading coincides with the flanking nucleosomes . . . . . . 83
3.16 No evidence of nucleosome eviction at the Mcm2-7 loading position . 84
3.17 Nucleosome occupancy explains protected fragments at Mcm2-7 loading position . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 85
3.18 Mcm2-7 loads either upstream or downstream of ORC . . . . . . . 87
3.19 Abf1p does not physically interact with adjacent nucleosomes . . . 88
3.20 Mcm2-7 loading at either flanking nucleosome can catalyze replication initiation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89

xii
List of Abbreviations

AAA+ ATPases associated with a variety of cellular activities
ACS ARS consensus sequence
APC/C Anaphase-promoting complex/cyclosome
ARS Autonomously replicating sequence
ATAC-seq Assay for transposase-accessible chromatin using sequencing
ATP Adenosine triphosphate
BrdU Bromodeoxyuridine
bp Base-pair
CDK Cyclin-dependent kinase
ChIP Chromatin immunoprecipitation
CMG Cdc45–Mcm2-7–GINS complex
cryo-EM Cryo-electron microscopy
DDK Dbf4-dependent kinase
DNase Deoxyribonuclease I
HU Hydroxyurea
Mcm2-7 Minichromosomal maintenance protein complex
MNase Micrococcal nuclease
NFR Nucleosome-free region
ORC Origin recognition complex
pre-IC Pre-initiation complex
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-LC</td>
<td>Pre-loading complex</td>
</tr>
<tr>
<td>pre-RC</td>
<td>Pre-replicative complex</td>
</tr>
<tr>
<td>S-CDK</td>
<td>S-phase cyclin-dependent kinase</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single-stranded DNA</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start site</td>
</tr>
<tr>
<td>UASg</td>
<td>Upstream activation sequence G</td>
</tr>
</tbody>
</table>
Acknowledgements

Completion of a dissertation requires guidance and support from many people, and I have been extremely fortunate to have received both from so many. I would first like to thank my advisor, David MacAlpine, for all of his guidance and mentorship during my graduate studies. He has been tremendously supportive, always available to talk about research or life, and has provided a great research environment. I would also like to thank my dissertation committee for taking the time and effort to assist in my research endeavors: Alex Hartemink, who served as my de facto co-advisor, was always willing to help, and provided exceptional insight on countless matters; Ryan Baugh, who helped tremendously with my transition to graduate school during my first rotation; Tom Petes, for thinking of innovative ways to extend my research to additional biological questions; and, Tim Reddy, for always proposing great research questions.

I entered the CBB program with limited previous computational experience, and I am indebted to the many people who helped me establish a solid foundation of computational biology. A special thanks to Matt Eaton, Joey Prinz, and David Corcoran for providing me with countless scripts and for always lending a helping hand with my computational analyses.

I would like to thank each current and former member of the MacAlpine Lab, Rachel Creager, Leyna DeNapoli, Queying Ding, Matt Eaton, Monica Gutierrez, Yulong Li, Yoav Lubelsky, Heather MacAlpine, Sara Powell, Joey Prinz, and Vinay
Tripuraneni, for providing an engaging laboratory environment. A special thanks to Heather MacAlpine for her dedication to the MNase experiments that yielded exceptional data for analysis. Thanks also to Don Fox and Sarah Goetz, along with their respective labs, for giving me great feedback during my lab meeting presentations.

My project was initiated by a fruitful collaboration with Steve Henikoff, and I am grateful for his generosity in sharing his data and protocols.

The Duke CBB program is filled with tremendous people, establishing a rich training experience that ensures its students success in the program. I would first like to thank each of the CBB directors, John Harer, Alex Hartemink, and Paul Magwene, and DGSs, Jeanette McCarthy, Steve Haase, and Scott Schmidler, for their service and dedication to the program. I would also like to thank Liz Labriola for all her work in ensuring that everything runs smoothly for the program. My interactions with all the graduate students within the CBB program have greatly enhanced my research, and in particular I’d like to thank Jianling Zhong, Kaixuan Luo, Michael Mayhew, and Anirudh Natarajan for their insights into my work.

Finally, but certainly not least, I'd like to thank my parents, Ro and Jeff Belsky, and sister, Janine Belsky, for their unwavering support, guidance, and love. They have provided me with incredible opportunities to pursue my goals, teaching and instilling in me values that have crafted the person I’ve become.
Introduction

Genomics has been revolutionized by the (nearly) completed sequenced genomes of many organisms. Even though genome sequences are typically depicted as linear strings, this simplified view neglects the three-dimensional chromatin configuration by omitting nucleosomes and other DNA-binding proteins. It is becoming increasingly clear that the binding locations of these factors play a substantial role in regulating many DNA-templated processes. A clear understanding of how the coordinated assembly of proteins collectively interprets the genome would provide an increased understanding of genomic processes and better guide the development of new therapeutics.

Recent genome-wide datasets have revealed many aspects of the chromatin structure, from the distribution of nucleosomes to the enrichment locations of transcription factors. However, to date, no assay can simultaneously resolve the positions of all DNA-binding factors. In Chapter 2 of this dissertation, I describe an extension of the micrococcal nuclease (MNase) digestion assay to query protein occupancy throughout the genome. I will demonstrate how differences in the widths of recovered DNA fragments distinguish between nucleosome occupancy and transcription
factor (TF) binding. Upon sequencing all DNA fragments remaining after an MNase digestion of *Saccharomyces cerevisiae* chromatin, we detected high-resolution TF footprints *in vivo* across the genome. I set out to comprehensively evaluate these fragment protections at previously identified transcription-factor-binding sites across the *S. cerevisiae* genome, and ask if the fragment protections reveal further insights into protein interactions with specific DNA sequences and nucleosomes. This work was performed as a collaboration with the Henikoff group, and was published in the *Proceedings of the National Academy of Sciences* (Henikoff et al. 2011). This is a powerful approach that, when used in conjunction with sequence annotations and chromatin immunoprecipitation (ChIP) data, can thoroughly profile the entire chromatin environment in one experiment.

In Chapter 3 of this dissertation, I apply this MNase digestion method to specifically profile the chromatin architecture surrounding replication origins. First, I identified the subset of potential replication origins bound by the origin recognition complex (ORC), the initial protein present at replication origins, by detection of a high-resolution *in vivo* protein footprint. I then surveyed changes in the chromatin architecture as additional proteins are recruited to the replication origin, culminating in the formation of the pre-replicative complex (pre-RC). My analysis revealed a dynamic chromatin structure at replication origins during pre-RC assembly, and linked several distinct chromatin configurations to replication function. This research was recently published in *Genes & Development* (Belsky et al. 2015).

Overall, my work in the development of a new technique to profile the chromatin structure has led to additional insights into the transcription and replication programs, and serves as the basis for future chromatin studies using this assay.
1.1 Chromatin

The linear sequence of DNA along each chromosome must be simultaneously compacted within the nucleus, yet remain accessible to permit genomic processes such as transcription, replication, and DNA repair. The interplay between these two seemingly contradictory goals is primarily mediated by nucleosomes, the fundamental unit of chromatin that wraps ~147 bp of DNA around a histone octamer (Finch et al. 1977). Although nucleosomes are regularly spaced throughout the genome, occupying 80% of all available DNA (Lee et al. 2007), certain regions are devoid of nucleosomes whereas other areas harbor higher-order nucleosome structures. Further, this compaction level can dramatically change depending on environmental context or, in the case of multicellular organisms, differentiation, leading to vastly different chromatin configurations along the same genomic sequence. Recent studies point to a myriad of different regulatory mechanisms that together can fine-tune and localize a preferential chromatin architecture. Nucleosomes can be distinguished by marking them with specific post-translational modifications or building them with different histone variants. Although the genome contains preferential nucleosome localization signals, these sequence determinants can be overwritten by trans-acting factors such as ATP-dependent chromatin remodelers. Together, these enable conserved and stereotypical nucleosome configurations at defined genomic elements such as transcription start sites (TSSs) and replication origins.

1.1.1 Nucleosome structure

Nucleosomes are the fundamental unit of chromatin, and consist of four distinct histones, H2A, H2B, H3, and H4 (Kornberg 1974; Kornberg and Thomas 1974), that are highly conserved throughout eukaryotes (Malik and Henikoff 2003). Each histone protein folds into a series of alpha helices, forming a histone-fold protein domain
that facilitates dimerization among the histones (Luger et al. 1997). The most stable histone-histone interaction is between H3 and H4, followed by H2A and H2B, forming two types of heterodimers. The histone octamer is formed by first stringing two sets of H3-H4 heterodimers together via H3:H3 interactions. Subsequent H4:H2B interactions then join two H2A-H2B heterodimers to the H4-H3:H3-H4 tetramer. When these four sets of heterodimers are brought together, they form a 5.5 nm long cylindrical structure spanning an 11 nm diameter (Finch et al. 1977).

Nucleosomes facilitate DNA compaction by stabilizing the negatively-charged phosphate DNA backbone. The DNA double-helix is wrapped ~1.75 times, or 147 bp, around the histone octamer core in a left-hand spiral (Finch et al. 1977). Sequential DNA-wrapped nucleosomes are separated by linker DNA, forming the primary chromatin structure known as the 10-nm fiber (Olins and Olins 1974). This primary structure can be further compacted into higher-order chromatin structures known as the 30-nm fiber. Two models have been proposed for the 30-nm fiber, either the solenoid (Finch and Klug 1976) or zigzag (Thoma et al. 1979) model, differing in whether the primary inter-nucleosome interactions occur between proximal or alternate nucleosomes. These higher-order configurations likely require the linker histone H1 to stabilize this high-density packing structure.

Although these higher-order chromatin structures efficiently compact the genome within the nucleus, they sterically inhibit other chromatin proteins from accessing the genome. Thus, compaction of the genome must be selectively regulated to allow these proteins access when required. A key step in deciphering chromatin regulation is understanding what determines the degree of nucleosome packing at specific genomic regions.

One way to locally regulate the chromatin structure is through post-translational modifications of histones. Each histone contains an unstructured N-terminal tail that contains many potential sites for post-translational modifications, such as methyla-
tion and acetylation (Millar and Grunstein 2006). These modifications can alter the interactions at the histone-DNA interface, or alternatively create binding sites for proteins harboring specific recognition domains (Strahl and Allis 2000). Many histone modifications have been mapped throughout the genome, and some modifications are associated with high transcription rates or are enriched relative to defined genomic elements like transcription start sites (Liu et al. 2005). While these modifications do not significantly impact how the DNA is wrapped around the histone octamer, they do affect the stability of the nucleosome and play a role in organizing higher-order chromatin.

In addition to post-translational modifications, nucleosomes are further distinguished across the genome by the inclusion of several histone variants that can substitute for the primary core histones. Like with post-translational modifications, these histone variants are typically localized to specific regions of the genome, giving these variant-containing nucleosomes unique properties to affect genomic processes. For example, nucleosomes surrounding transcription start sites tend to substitute H2A with the histone variant H2A.z (Guillemette et al. 2005; Albert et al. 2007). At centromeres, a special H3 variant, CENP-A, replaces H3 and modifies how the DNA is wrapped around the histone octamer (Camahort et al. 2009). Thus, histone variants provide another way to differentiate nucleosomes across the genome.

1.1.2 Nucleosome positioning

Nucleosomes are the gatekeepers for protein access to the genome. DNA that is wrapped within a nucleosome is much less accessible for protein binding compared to the linker DNA between nucleosomes. Thus, a key question is understanding the determinants of nucleosome positioning throughout the genome.

Nucleosome positioning is partially encoded by the DNA sequence. Early studies showed that wrapping the DNA double helix around the histone octamer requires sig-
significant bending (Zhurkin et al. 1979). Certain DNA sequence features with narrow
minor grooves, such as 10-bp repeats of AA/TT/AT dinucleotides, promote bending
of the DNA around histone octamers (Satchwell et al. 1986). Since DNA is twisted
around the histone octamer, the double helix is more thermodynamically stable if
the narrow minor grooves of these dinucleotides face the nucleosome surface, resulting
in a preferred rotational positioning configuration (Drew and Travers 1985). In
contrast, stretches of consecutive A:T sequences typically prevent the formation of
nucleosomes by creating unfavorable energetic conditions for DNA bending (Satch-
well et al. 1986).

Several groups developed statistical models based on these sequence features in
an effort to predict nucleosome locations throughout the genome (Ioshikhes et al.
2006; Segal et al. 2006; Kaplan et al. 2009). Although these models could recapit-
ulate several aggregate chromatin features, such as those around transcription start
sites, there was only a 50% agreement with in vivo nucleosome positions, suggesting
additional determinants influence nucleosome locations.

In addition to sequence, nucleosome positioning is impacted by the presence of
other DNA-binding proteins in the local proximity, known as the “barrier” model (Ko-
rnberg 1981; Kornberg and Stryer 1988). If a highly localized chromatin-binding pro-
tein demarcates a fixed position, the nucleosomes surrounding this “barrier” must
be located relative to this boundary element, taking into account steric hindrances
and minimum linker length requirements between nucleosomes. Importantly, the
effect of this “barrier” element can be propagated along the nucleosome array, so
that restricting the proximal nucleosome position impacts the location of more distal
nucleosomes. In this case, the presence of the barrier element supercedes sequence
preferences for nucleosome positioning.

Nucleosome positioning is also impacted by specialized proteins called ATP-
dependent chromatin remodelers. These are typically multi-subunit protein com-

plexes that, through the utilization of ATP, actively shift or remove nucleosomes. There are four main families of ATP-dependent chromatin remodeler complexes, SWI/SNF, ISWI, CHD, and INO80 (Clapier and Cairns 2009). Their activities are typically very widely distributed throughout the genome, yet they can be localized by interactions with certain histone modifications. For example, SWI/SNF contains a bromodomain that recognizes acetylated lysines (Hassan et al. 2002), and the chromodomain of CHD interacts with methylated lysines (Pray-Grant et al. 2005). By shifting nucleosomes, ATP-dependent chromatin remodelers also play a significant role in transitioning between transcriptional states.

1.1.3 Nucleosome positioning at the TSS

Recent advances in genomics have revealed the nucleosome distribution throughout the genome. An important question is understanding the role of nucleosome positioning in regulating genomic processes such as transcription. Aligning these nucleosome maps on transcription start sites has revealed a canonical distribution of well-positioned nucleosomes surrounding promoters.

An important general feature of promoters is a well-defined nucleosome-free region (NFR) upstream of the transcription start site. Across yeast promoters, the NFR has an average width of 140 bp and is flanked by two well-positioned nucleosomes, known as the -1 and +1 nucleosome (reflecting the first upstream and first downstream nucleosome relative to the TSS, respectively). By keeping the promoter region free of nucleosomes, loading of the transcriptional machinery is more easily facilitated (Knezetic and Luse 1986).

Specific sequence features play a role in keeping the promoter regions nucleosome-depleted. First, the NFR typically contains poly(A:T) tracts that prevent nucleosome encroachment over the promoter region (Yuan et al. 2005). Further support for the role of these exclusionary sequences comes from datasets evaluating favorable
nucleosome locations. Even when nucleosomes are permitted to freely sample \textit{in vitro} their preferred locations in the absence of other proteins, promoter regions still remain devoid of nucleosomes (Kaplan \textit{et al.} 2009; Zhang \textit{et al.} 2009). In contrast, the upstream and downstream regions flanking the NFR contain preferred nucleosome positioning signals in the form of 10-bp repeats of AA/TT dinucleotides, contributing to the positioning of the -1 and +1 nucleosomes (Ioshikhes \textit{et al.} 2006; Mavrich \textit{et al.} 2008a).

In addition to the direct role of nucleosome positioning sequence features, transcription factor binding contributes to keeping the promoter void of nucleosomes. The NFR is particularly enriched in transcription-factor–binding sites (Harbison \textit{et al.} 2004), and TFs can serve as boundary elements to position nucleosomes. For example, nucleosomes are precisely positioned around Abf1 and Reb1 TF-binding sites. This is dependent on Abf1 and Reb1 occupancy, as nucleosomes show no consensus positioning around these sites when histones are reassembled onto DNA \textit{in vitro} (Kaplan \textit{et al.} 2009; Zhang \textit{et al.} 2009). Similarly, ablation of Abf1 or Reb1 DNA-binding with temperature-sensitive mutants leads to nucleosome accumulation over their respective binding sites (Badis \textit{et al.} 2008).

Promoters can also simultaneously utilize several features to establish the NFR. Multiple TFs may bind synergistically to a specific promoter, and their individual binding events may collectively help to exclude nucleosomes (Bai \textit{et al.} 2011). This combinatorial binding model likely occurs at genes harboring weaker nucleosome positioning signals, as these promoters contain additional TF-binding sites (Mavrich \textit{et al.} 2008a). TF-binding sites can also occur in promoters containing a nucleosome exclusionary poly(A:T) tract (Lascaris \textit{et al.} 2000). This cooperativity can be especially effective at creating NFRs; transferring a short promoter sequence with a poly(A:T) tract and Reb1-binding site to an ectopic location induces nucleosome depletion (Raisner \textit{et al.} 2005).
In addition to the precise positioning of the -1 and +1 nucleosomes, nucleosomes are also phased throughout the gene body. These nucleosomes are typically separated by an ~18-bp linker, and the conserved spacing dissipates as the nucleosome array proceeds towards the 3’ end of the gene. How is this nucleosome positioning maintained throughout the gene body? Sequences favorable for nucleosome formation are found underlying the first few nucleosomes in the array, albeit at lower levels than the +1 nucleosome (Mavrich et al. 2008a). If these nucleosomes are primarily positioned by sequence, this nucleosome pattern should be prevalent in the in vitro nucleosome maps; however, the only dominant feature at transcription start sites is nucleosome depletion over the promoter (Kaplan et al. 2009; Zhang et al. 2009). No consistent nucleosome phasing was observed over the gene body.

One possibility is that the nucleosome exclusionary sequence is not sufficient to precisely position the +1 nucleosome. Instead, the presence of transcription factors and/or the transcription machinery may be necessary to direct the +1 nucleosome into its conserved location. Once the +1 nucleosome is positioned, stacking the remaining nucleosomes over the gene body against this barrier would result in consensus positioning (Mavrich et al. 2008a). Thus, repeating the in vitro nucleosome reconstitution experiment in the presence of chromatin proteins should result in nucleosome phasing. However, inclusion of a whole cell extract did not alter the in vitro nucleosome positions (Zhang et al. 2011), suggesting that nucleosome phasing requires an additional activity. Interestingly, by supplementing the whole cell extract with ATP, substantial nucleosome phasing is now apparent throughout the gene body, suggesting that ATP-dependent chromatin remodelers are responsible for establishing this chromatin configuration (Zhang et al. 2011). In support of this idea, several chromatin remodeler immunoprecipitation experiments have shown that ATP-dependent chromatin remodelers are targeted to specific nucleosome positions surrounding the TSS (Whitehouse et al. 2007; Yen et al. 2012). Deletion of chro-
matin remodelers also results in a loss of nucleosome phasing (Gkikopoulos et al. 2011). Most of these changes did not affect the positions of the -1 and +1 nucleosomes; instead, nucleosome phasing within the gene body and further upstream were lost.

Even though genes have varying levels of transcription, the overall nucleosome structure does not differ substantially when comparing low-transcribed genes to high-transcribed genes. The most pronounced difference is the nucleosome occupancy over the gene body, which is less apparent in highly transcribed genes yet still demonstrates phasing (Yuan et al. 2005; Mavrich et al. 2008a). In particular, the occupancy of the +1 nucleosome is not dependent on the transcription rate (Mavrich et al. 2008a; Shivaswamy et al. 2008), yet the DNA wrapped around these histone octamers must be temporarily relaxed to permit passage of the RNA polymerase (Bondarenko et al. 2006). As DNA spontaneously dissociates from the nucleosome (Li et al. 2005), the RNA polymerase can gain access to the DNA when it encounters this unwrapped state. The FACT complex assists in displacing one H2A-H2B dimer from the histone octamer (Kireeva et al. 2002; Belotserkovskaya et al. 2003). After the RNA polymerase has passed, FACT also assists in reforming the histone octamer by localizing the displaced H2A-H2B dimer near the disrupted nucleosome (Schwabish and Struhl 2004).

1.1.4 Dynamic chromatin structure affects transcriptional regulation

Yeast promoters can be subdivided into two transcription classes, constitutive and inducible. Constitutively transcribed genes tend to lack a TATA-box and contain the promoter architectures described above (Fig. 1.1). In contrast, inducible genes harbor a TATA-box, have a much less well-defined NFR (Albert et al. 2007; Mavrich et al. 2008a; Shivaswamy et al. 2008), and are likely more dependent on chromatin remodelers to activate transcription (Weiner et al. 2012). Inducible genes regulate
transcription by alternating between restrictive and permissive chromatin structures. The *GAL1-10* promoter region provides a well-studied example linking chromatin dynamics to gene induction on the basis of sugar availability. In the absence of galactose, the *GAL1-10* promoter region is covered by an array of nucleosomes (Lohr 1984); however, addition of galactose results in a significant reconfiguration of the chromatin structure (Cavalli and Thoma 1993). The presence of galactose frees the Gal4 transcription factor from Gal80 repression, allowing Gal4 localization to four binding sites within the promoter NFR, known as the *UASg*. Gal4 binding can then recruit the transcriptional machinery to begin transcription of the *GAL1* and *GAL10* genes. However, what enables Gal4 to bind to the *UASg*, which is initially occluded by a nucleosome? Although Gal4 is capable of outcompeting a nucleosome for access to this site, Gal4 binding is facilitated by the constitutive presence of the RSC chro-
matin remodeler at the nucleosome occluding the $UAS_g$ (Floer et al. 2010). RSC keeps the nucleosome at this location in a partially unwrapped state, which allows Gal4 easier access to this location (Floer et al. 2010). Gal4 binding also attracts the SWI/SNF chromatin remodeling complex to promote further nucleosome depletion (Bryant et al. 2008), establishing an NFR to enable transcription. Even though $GAL1$ and $GAL10$ can be upregulated in the absence of RSC and SWI/SNF, the dynamics of nucleosome turnover are much slower. This demonstrates a cooperativity between chromatin remodelers and transcription factors as a means for upregulating transcription.

A similar regulatory chromatin structure is present at the $PHO5$ promoter, which becomes active only in the absence of phosphate. Similar to the $GAL1-10$ promoter, nucleosomes cover the $PHO5$ promoter in the “off” state (Almer and Hörz 1986). The $PHO5$ promoter region contains two binding sites for the transcriptional activator Pho4. One binding site is accessible in the linker region between two nucleosomes, whereas the other is occluded by a nucleosome (Svaren and Hörz 1997). Sufficient levels of phosphate result in the phosphorylation of the Pho4 protein, preventing its binding to the accessible site (Kaffman et al. 1994). In the absence of phosphate, Pho4 repression is relieved, allowing Pho4 localization to the accessible binding site in the linker region. This binding causes a vast restructuring of the chromatin architecture (Almer et al. 1986; Boeger et al. 2003), including loss of the four nucleosomes flanking the Pho4 linker region binding site. As in regulation at the $GAL1-10$ promoter, the temporal control of this nucleosome depletion is facilitated by the chromatin remodelers SWI/SNF (Gaudreau et al. 1997) and INO80 (Barbaric et al. 2007), yet neither of these are required. In addition, the histone acetylases NuA4 (Nourani et al. 2004) and Gcn5 (Gregory et al. 1998; Barbaric et al. 2001) also promote nucleosome removal. Thus, multiple chromatin regulators play a role in remodeling the $PHO5$ promoter.
The principles elucidated from the *GAL1-10* and *PHO5* promoters are also found more globally when evaluating stress-responsive genes. When yeast undergo a heat shock stress, transcription is inhibited at ribosomal genes in part by nucleosome occlusion over the promoter region. This nucleosome occupancy likely impedes RNA Polymerase II binding and transcription initiation (Venters and Pugh 2009). In contrast, heat-shock responsive factors, such as Msn4, become active and can compete with nucleosomes for access to their binding sites, resulting in nucleosome eviction at upregulated promoters (Shivaswamy *et al.* 2008). These studies demonstrate how interconverting chromatin configurations can dramatically affect transcription.

1.2 Genomic approaches for studying chromatin

A number of techniques have been developed to uncover the spatial locations of DNA-binding proteins throughout the genome. Although chromatin profiling was originally restricted to individual DNA templates, chromatin profiling has been extended genome-wide with recent advances in microarray technology and next-generation sequencing. Even though each individual assay has its own limitations, combining results across techniques yields a comprehensive evaluation of the chromatin occupancy landscape.

1.2.1 Chromatin immunoprecipitation (ChIP)

The “gold standard” assay for profiling DNA-binding proteins is chromatin immunoprecipitation (ChIP) (Fig. 1.2) (Gilmour and Lis 1984; Ren *et al.* 2000; Lieb *et al.* 2001; Lee *et al.* 2002; Barski *et al.* 2007; Johnson *et al.* 2007; Mikkelsen *et al.* 2007; Albert *et al.* 2007). In this experimental technique, the genomic binding locations of a candidate protein are queried by first creating a covalent, static interaction between the factor and its binding sequence via formaldehyde cross-linking (Jackson 1978), ultraviolet light (Becker and Wang 1984), or dimethyl sulfate (Levina *et al.*
Figure 1.2: ChIP-seq schematic. DNA regions bound by a protein of interest (green oval) are specifically recovered by an antibody (black). Sequencing library adapters (orange) are ligated to each end of the recovered fragments, and high-throughput sequencing yields an enrichment of reads in the protein-binding location (green rectangles).

Next, the genome is randomly fragmented into smaller segments, typically \( \sim 100-400 \) bp, by utilizing restriction enzymes or sonication. This chromatin extract now contains all DNA-binding proteins locally associated with their interacting genomic sequences. Next, an individual candidate factor of interest is specifically enriched by utilizing an antibody, which preferentially recognizes an epitope only found on the candidate factor. The antibody allows for the selective recovery of the
DNA-binding factor crosslinked to its associated sequences via a process called immunoprecipitation (Gilmour and Lis 1984). Reversing the DNA crosslinks with SDS and removing the proteins with proteinase K then allows recovery of the interacting DNA sequences. The resulting DNA can then be analyzed by either Southern Blot, PCR, tiling array (ChIP-chip), or sequencing (ChIP-seq) to identify the enriched genomic regions.

Although ChIP-chip and ChIP-seq can reveal the spatial locations of specific factors, the random fragmentation step yields DNA sequencing reads that typically span several hundred base-pairs, much wider than the predicted DNA-binding footprint. To more finely pinpoint the precise DNA-binding location, fragmented reads are further trimmed to the protein footprint by digestion with lambda exonuclease. Applying this approach, termed ChIP-exo, to transcription factors revealed two narrow signal peaks surrounding the DNA-binding footprint, confirming that site-specific factors consistently bind to the same location across a population of cells (Rhee and Pugh 2011). ChIP-exo has also been extended to chromatin remodelers (Yen et al. 2012), histones (Rhee et al. 2014), and components of the transcription pre-initiation complex (Rhee and Pugh 2012), further refining their binding distributions throughout the genome, though not to the resolution of site-specific DNA-binding factors.

Despite the high resolution and specificity achieved with ChIP, there are several drawbacks to this approach. First, although the advantage of immunoprecipitation is its specificity, this limits the throughput of ChIP assays to profiling one factor per experiment. Thus, scaling this assay either “horizontally” to profile all DNA-binding factors or “vertically” to profile one specific DNA-binding factor in a variety of conditions or across multiple time points becomes time consuming, labor-intensive, and expensive. The premise of chromatin immunoprecipitation is that an antibody recognizing a specific epitope on the protein is readily available. However, specific
antibodies do not exist for every DNA-binding protein, or the antibody specificity is suboptimal (Teytelman et al. 2013; Park et al. 2013). Further, appending recognizable tags like His or FLAG may either hinder the protein function or require novel genome engineering techniques to ensure that the protein concentration remains at canonical in vivo levels. All of these concerns necessitate alternative genome-profiling approaches to both validate the results obtained from ChIP and increase the throughput beyond profiling one factor per assay.

1.2.2 Chromatin profiling by DNA digestion

As an alternative to ChIP, several genomic methods have been developed that profile the entire chromatin architecture without immunoprecipitation. Instead of selectively enriching for one factor as in ChIP, these assays agnostically measure protein occupancy across the genome. These techniques involve non-specifically digesting the genome in locations unoccupied by DNA-binding proteins. Then, analyzing the resulting DNA digestion patterns yields the spatial occupancy distribution of DNA-binding proteins.

Deoxyribonuclease I (DNase)

The most widely-used enzyme in this class is deoxyribonuclease I (DNase). DNase is an endonuclease, producing individual “nicks” non-specifically throughout the genome (McCarty 1946). If DNase is subjected to a “naked” DNA template without any DNA-binding proteins, it will convert the DNA polymer into small oligonucleotides. Many eukaryotes may utilize DNase to eliminate unwanted foreign DNA in the cell (Samejima and Earnshaw 2005).

However, if DNase encounters a chromatinized template, a competition ensues between DNA-binding proteins and DNase for preferential access to specific DNA sites (Galas and Schmitz 1978). By titrating the DNase concentration and mod-
Figure 1.3: DNase schematic. Open chromatin regions are susceptible to DNase cleavage (purple arrows). Recovery of DNA fragments in these regions, coupled with ligation of sequencing adapters (orange), yields selective read enrichment via Illumina sequencing (purple rectangles). These DNase hypersensitive sites typically occur at intergenic, nucleosome-free regions where DNA-binding proteins (green circle and blue rectangle) are localized.

Regulating its digestion time, optimization of the DNase assay ensures that genomic positions not bound by a protein are preferentially cleaved by DNase. In contrast, genomic regions harboring a DNA-binding protein remain protected from DNase digestion (Crawford et al. 2006; Sabo et al. 2006; Hesselberth et al. 2009). In practice, because DNase is ~31 kDa (Price et al. 1969), steric hindrance reduces the ability of DNase to digest within the short linker region between nucleosomes. Thus, only open chromatin regions, genomic locations devoid of nucleosomes such as gene promoters, are susceptible to DNase cleavage (Fig. 1.3). Several algorithms have been developed to evaluate transcription-factor occupancy at candidate binding motifs given the local DNase digestion pattern (Hesselberth et al. 2009; Chen et al. 2010; Pique-Regi et al. 2011; Luo and Hartemink 2013; Yardimci et al. 2014). An observed TF foot-
print would result in “hypersensitive” DNase digestion proximal to the TF-binding site and protection from DNase cleavage over the TF-bound motif sequence.

DNase footprinting helps alleviate the major drawback from ChIP experiments in that it can agnostically profile all TF-binding events across the genome. However, DNase can only reveal a subset of chromatin proteins, as nucleosome locations remain obscured. In conjunction with TF binding, nucleosome positioning plays a substantial role in transcriptional regulation (Bai and Morozov 2010), so DNase data has to be complemented with alternative nucleosome-mapping approaches.

Assay for transposase-accessible chromatin using sequencing (ATAC-seq)

Although DNase digestion can effectively probe the chromatin structure of open regulatory regions, the protocol requires millions of cells and many DNA-processing steps. In contrast, a complementary approach, termed transposase-accessible chromatin using sequencing (ATAC-seq), overcomes these obstacles by utilizing fewer cells, on the order of hundreds to thousands, and limiting the number of sample preparation steps (Buenrostro et al. 2013). Instead of cleaving open DNA regions with an endonuclease, ATAC-seq utilizes the transposase Tn5 to insert Illumina sequencing adapters directly into accessible genomic regions. DNA regions not bound by proteins are more susceptible to transposition by Tn5 (Gangadharan et al. 2010), preferentially tagging open accessible regions with sequencing primers. This approach also successfully interrogates DNA accessibility across the genome.

Micrococcal nuclease (MNase)

Nucleosome positioning has classically been profiled with the endo-exonuclease micrococcal nuclease (MNase), a 5'-phosphodiesterase that mediates the metabolic processing of nucleic acids in Staphylococcus aureus (Cunningham et al. 1956; Catlin and Cunningham 1958; Reddi 1960; Tucker et al. 1978). The smaller size of micrococcal
nuclease (~12 kDa) (Alexander *et al.* 1961) in conjunction with its exonuclease activity makes it more amenable to digesting the linker region between nucleosomes, allowing for the resolution of a single nucleosome (Fig. 1.4) (Noll 1974; Noll *et al.* 1975; Kornberg 1977; Cockell *et al.* 1983). Initially, nucleosomes at select genomic loci were interrogated with MNase (Struhl 1982; Thoma *et al.* 1984). With the advances of higher resolution platforms, nucleosome profiling was extended genome-wide first with tiling arrays (Sekinger *et al.* 2005; Yuan *et al.* 2005), and later with next-generation sequencing (Albert *et al.* 2007; Mavrich *et al.* 2008a,b).
Even though MNase historically has been applied exclusively to nucleosome mapping, several groups demonstrated that MNase could also resolve smaller protein footprints, such as those found at gene promoters (Ganter et al. 1993; Gerlach et al. 1995; Teng et al. 2001). Detailed analysis of the MNase digestion preparations revealed a subset of DNA fragments that were significantly smaller than the ~150-bp width expected of nucleosomes. These smaller fragments, termed subnucleosome particles, were hypothesized to represent either unwrapped nucleosomes (Floer et al. 2010) or DNA-binding proteins in promoter regions (Weber et al. 2010). In Chapter 2 of this dissertation, I will explore the role of these smaller fragments in depth at TF-binding sites across the *S. cerevisiae* genome.

1.3 DNA Replication

DNA replication is a fundamental, highly regulated genomic process that ensures faithful and complete inheritance of genetic information during each cell cycle. Errors in DNA replication can induce genomic instability, resulting in gene mutations, deletions, or amplifications, or larger-scale genome abnormalities such as chromosomal translocations, addition, or loss. Each of these disruptions to genome integrity can lead to cell cycle dysregulation, commencing the cascade of events leading to uncontrolled cell growth and cancer (Lengauer et al. 1998; Shen 2011). For this reason, each step of the DNA replication program is restricted to a specific period during the cell cycle, and is tightly controlled by several redundant mechanisms, limiting the possibility of detrimental errors.

1.3.1 Pre-replicative complex (pre-RC) formation

Before DNA polymerases can commence duplication of the genome, the replication machinery must be recruited to DNA replication initiation sites, known as replication origins. The first protein at these locations is the origin recognition complex.
(ORC) (Bell and Stillman 1992), a hetero-hexameric protein highly conserved among all eukaryotes (Gavin et al. 1995; Tugal et al. 1998). Five of the six ORC subunits (Orc1p-5p) contain the AAA\(^+\) (ATPases associated with a variety of cellular activities) protein domain (Neuwald et al. 1999), and each of these subunits also contains a winged-helix domain that facilitates contact with the DNA (Clarey et al. 2006; Gaudier et al. 2007). These ORC subunits are further organized into two higher-order structures (Chen et al. 2008). One domain contains Orc1p, Orc4p, and Orc5p, which interfaces with the separate Orc2p-Orc3p complex. The remaining ORC subunit, Orc6p, does not contact the DNA directly and primarily interacts with the Orc2p-Orc3p domain.

ORC binding establishes a 48-bp footprint (Bell and Stillman 1992; Santocanale and Diffley 1996; Speck et al. 2005) and is mediated by individual Orc1p-5p subunit interactions along a 30-bp region of the replication origin DNA (Lee and Bell 1997; Chastain II et al. 2004). The specificity of ORC-DNA interactions is regulated by ATP binding to Orc1p. ATP is required both for ORC binding to DNA, as well as constitutive residency of ORC on the DNA; non-specific ORC-DNA interactions facilitate ATP hydrolysis and ORC dissociation from the chromatin (Klemm et al. 1997). ORC assembly on the DNA primes the replication origin for loading additional replication initiation proteins.

ORC first recruits Cdc6 to the replication origin (Fig. 1.5). Cdc6 was initially discovered in a screen for cell cycle mutants (Hartwell et al. 1973), and further genetic studies with ORC mutants revealed a significant role in DNA replication initiation (Liang et al. 1995). As a member of the AAA\(^+\) protein family (Liu et al. 2000), Cdc6 can also bind ATP, which is required to form a functional ORC-Cdc6 complex (Speck et al. 2005). Cdc6 binding induces a structural rearrangement of the ORC subunits, which further increases the specificity of ORC for replication origin DNA (Mizushima et al. 2000; Speck and Stillman 2007) and renders significant
binding sites on Orc6p accessible for other replication proteins (Sun et al. 2012). Cdc6 binding also extends the ORC footprint by 30 bp (Santocanale and Diffley 1996; Speck et al. 2005).

The primary role of the ORC-Cdc6 complex is to recruit the Mcm2-7 helicase to DNA; once Mcm2-7 is loaded onto the DNA, both ORC and Cdc6 are no longer needed for replication to proceed (Donovan et al. 1997). The Mcm2-7 is the replica-
tive helicase, and like ORC, is a hetero-hexameric protein composed of six distinct subunits (Mcm2p-7p) (Bochman and Schwacha 2009). These six subunits are organized into a ring surrounding a central cavity, which accommodates the DNA after stable loading of the Mcm2-7 hexamer. Each of the Mcm subunits contains a AAA+ domain, and ATP hydrolysis contributes to helicase unwinding of the DNA.

Mcm2-7 requires a binding partner, Cdt1, to associate with replication origins, and interactions between Cdt1 and Mcm6p (Yanagi et al. 2002; You and Masai 2008; Jee et al. 2010; Wu et al. 2012) form a Cdt1–Mcm2-7 complex that is imported into the nucleus (Tanaka and Diffley 2002). Although single Mcm2-7 hexamers are recruited to the replication origin, Mcm2-7 will ultimately be loaded as a head-to-head double hexamer (Remus et al. 2009; Evrin et al. 2009). Two Cdt1-binding sites on Orc6p (Chen et al. 2007; Takara and Bell 2011), along with an interaction between Mcm3p and ORC-Cdc6 (Frigola et al. 2013), bring the Cdt1–Mcm2-7 complex to the replication origin.

Recruitment of Cdt1–Mcm2-7 to the replication origin results in an intermediate ORC/Cdc6/Cdt1/Mcm2-7 (OCCM) structure (Sun et al. 2013); however, additional steps are required for Mcm2-7 to stably encircle the DNA. First, the Mcm2-7 ring has to be opened to allow DNA access to the central channel. The Mcm2-7 ring is opened at the interface of the Mcm2p and Mcm5p subunits (Costa et al. 2011), yet it remains unknown if this occurs when the Cdt1–Mcm2-7 complex reaches the replication origin or is initiated by another active process. Cdc6 ATP hydrolysis initiates the first step in loading Mcm2-7 by closing the Mcm2-7 ring around double-stranded DNA and facilitating Cdt1 dissociation from the replication origin (Randell et al. 2006). Next, the C terminus of Mcm3p catalyzes ORC ATP hydrolysis (Fernández-Cid et al. 2013), releasing both Cdc6 and Mcm2-7 from their respective associations with ORC. ORC can now subsequently repeat these pre-RC formation steps to load another Mcm2-7 (Bowers et al. 2004).
Although many of the steps involved in Mcm2-7 loading have been investigated, questions still remain about how a head-to-head Mcm2-7 double hexamer is formed. Even though Orc6p can accommodate two Cdt1–Mcm2-7 complexes simultaneously, in vitro experiments suggest that one Mcm2-7 hexamer is initially loaded, followed by recruitment of a second Mcm2-7 hexamer (Fernández-Cid et al. 2013). Further, it is unclear how the second Mcm2-7 hexamer is loaded in an opposite orientation relative to the first hexamer.

1.3.2 Origin activation

Even though Mcm2-7 can passively translocate along the DNA once it has been stably loaded (Evrin et al. 2009; Remus et al. 2009), Mcm2-7 cannot unwind DNA without the presence of two additional factors, Cdc45 (Zou et al. 1997) and the GINS complex (Takayama et al. 2003). Together, Cdc45, Mcm2-7, and GINS form the CMG, which serves as the active motor unwinding DNA ahead of the replication fork (Gambus et al. 2006; Moyer et al. 2006; Ilves et al. 2010). CMG formation requires the activities of both the Dbf4-dependent kinase (DDK) and the S-phase cyclin-dependent kinase (S-CDK) (Labib 2010).

DDK activity increases at the conclusion of G1, and phosphorylation of the Mcm4p and Mcm6p subunits primes Mcm2-7 for the arrival of Cdc45 (Randell et al. 2010; Sheu and Stillman 2010). Cdc45 first interacts with the Sld3-Sld7 complex, and the Sld3 interaction with Mcm2-7 brings Cdc45 to the replication origin (Heller et al. 2011; Kamimura et al. 2001). Once S phase commences, S-CDK phosphorylates Sld2, catalyzing Sld2 association with the scaffold protein Dpb11. Dpb11 also associates with GINS and the leading-strand replicative polymerase ϵ forming a pre-loading complex (pre-LC) before reaching the replication origin (Muramatsu et al. 2010). S-CDK also phosphorylates Sld3 at the replication origin (Zegerman and Diffley 2010), and this Sld3 phosphorylation promotes an interaction with Dpb11,
bringing the pre-LC to the replication origin to form the pre-initiation complex (pre-IC).

Formation of the CMG leads to the dissociation of Sld2, Sld3, Sld7, and Dpb11 from the replication origin, and Mcm2-7 transitions from encircling double-stranded DNA to single-stranded DNA (Fu et al. 2011). In addition to polymerase ϵ, polymerases α and δ are subsequently recruited to the replication origin, serving as the primase and lagging-strand polymerase, respectively. Bidirectional replication can now commence from the replication origins, leading to genome duplication.

1.3.3 Regulation of pre-RC assembly and origin activation

In order to ensure that replication occurs once and only once per cell cycle, pre-RC formation and origin activation are limited to distinct regions of the cell cycle. This is primarily regulated by the actions of two complexes, cyclin-dependent kinases (CDKs) and the anaphase-promoting complex/cyclosome (APC/C) (Siddiqui et al. 2013). Both of these complexes require additional regulatory subunits to respectively direct CDK phosphorylation and APC/C ubiquitination to specific targets. CDK utilizes three types of regulatory subunits, collectively known as cyclins; availability of each cyclin class is strictly regulated to ensure activity only during specific periods of the cell cycle (Malumbres 2014). Likewise, APC/C requires regulatory adaptors such as Cdh1 or Cdc20 (Peters 2006). CDK regulates APC/C and vice-versa, leading to distinct oscillations of CDK and APC/C activity throughout the cell cycle.

During pre-RC assembly in G1, pre-IC formation and origin activation are prevented by low CDK activity. This is achieved in part by APC/C-mediated ubiquitination and subsequent degradation of S-phase cyclins (Shirayama et al. 1999). APC/C also ubiquitinates the regulatory subunit Dbf4, eliminating DDK activity during G1 (Oshiro et al. 1999; Weinreich and Stillman 1999; Ferreira et al. 2000). As both CDK and DDK activities are necessary for pre-IC formation, origin acti-
vation cannot occur during G1. Toward the end of G1, CDK activity increases by transcription of the G1-cyclin Cln2 (Dirick et al. 1995). Cln2-CDK phosphorylates the APC/C regulatory adaptor Cdh1, eliminating APC/C function and relieving APC/C repression of CDK activity (Zachariae et al. 1998). S-phase cyclins are now available to complement with CDK to promote pre-IC formation.

A number of regulatory mechanisms ensure that pre-RC formation cannot occur during S phase. Even though ORC remains constitutively bound at replication origins throughout the cell cycle in S. cerevisiae (Rowley et al. 1995; Liang and Stillman 1997; Fujita et al. 1998), CDK phosphorylation of the Orc2p and Orc6p subunits prevents Cdt1 recruitment to the replication origin (Nguyen et al. 2001). Cdc6 is targeted for degradation during S phase after CDK phosphorylation (Drury et al. 1997). In addition, Mcm2-7 is sequestered from the nucleus by CDK phosphorylation of Mcm3p (Labib et al. 1999; Liku et al. 2005). These mechanisms collectively ensure that Mcm2-7 cannot be actively loaded following origin activation, providing protection against re-replication.

1.3.4 Replication origins

DNA replication is spatially regulated by limiting initiation to defined genomic locations. The simplest method for controlling DNA replication would specify only one initiation location. Prokaryotes, with their small, circular genomes, successfully perform DNA replication in this manner; however, as eukaryotic chromosomes are orders of magnitude larger than prokaryotes, it would be infeasible to duplicate an entire chromosome with only one DNA replication origin. To alleviate this problem, eukaryotes distribute replication origins throughout each chromosome (every ~30 kb in yeast (Petes and Williamson 1975), every ~100 kb in higher eukaryotes (Yurov and Liapunova 1977; Besnard et al. 2012)).

DNA replication origins are most clearly defined in S. cerevisiae, and were initially
identified by isolating genomic segments responsible for plasmid survival (Stinchcomb et al. 1979). These specific regions are termed autonomously replicating sequences (ARSs) and represent 400–800 bp sequences that likely contain an initiator element. Sequence analysis revealed an 11-bp motif sequence, the ARS consensus sequence (ACS), that is conserved throughout these regions (Broach et al. 1983). Even though in *S. cerevisiae* the ACS is required for ORC binding and subsequent replication initiation (Bell and Stillman 1992), the presence of the ACS is not sufficient to specify a replication origin, as over 10,000 high quality ACS motif matches exist throughout the yeast genome (Breier et al. 2004). Further, not all ARS elements are functional *in vivo* (Dubey et al. 1991), suggesting that other factors contribute to defining replication origins.

Several additional sequence features are present at replication origins that contribute to enhanced replication function, though are not essential like the ACS. These sequence elements were initially identified at *ARS1* (Celniker et al. 1984; Marahrens and Stillman 1992), and although they are present at additional origins, they are much less conserved than the ACS. The B domain is located downstream of the ACS and consists of three distinct sequence elements. The B1 element is located most proximal to the ACS, and provides additional contact points for ORC (Lee and Bell 1997; Chastain II et al. 2004). Further downstream, the B2 element likely provides a preferred binding location for Mcm2-7. The most distal downstream element is B3, which serves as the binding site for Abf1. The C domain is located upstream of the ACS, and may specify the replication initiation location, as well as harbor binding sites for additional transcription factors. Importantly, active replication origins vary in their inclusion and spatial locations of these sequence elements.
1.3.5 Genome-wide mapping of replication origins

Replication origins have been mapped genome-wide in *S. cerevisiae* utilizing different replication properties. The binding distribution of ORC and Mcm2-7 have been revealed by chromatin immunoprecipitation, first coupled to tiling arrays (Wyrick *et al.* 2001; Xu *et al.* 2006) and more recently profiled by next-generation sequencing (Eaton *et al.* 2010). Despite the improved resolution of sequencing, the ChIP signal typically spans several hundred base pairs, requiring motif discovery to identify conserved sequences across the enriched regions. These experiments recovered the ACS as a common motif at replication origins (Xu *et al.* 2006; Eaton *et al.* 2010).

The number of replication origins identified by ChIP varies depending on the experiment and analysis thresholds. ORC and Mcm2-7 ChIP-chip experiments revealed 429 (Wyrick *et al.* 2001) and 349 (Xu *et al.* 2006) replication origins, whereas an ORC ChIP-seq experiment recovered 253 high-confidence sites (Eaton *et al.* 2010). Although complete concordance would be expected for ORC and Mcm2-7 ChIP binding sites, ChIP of Mcm2-7 has consistently been more efficient than ORC. As ORC binding is required for Mcm2-7 loading, this is likely a consequence of antibody epitope accessibility or the presence of at least two Mcm2-7 helicases at each replication origin.

Although ChIP-based approaches identify the locations of replication complex formation, they do not reveal information on replication origin function. Replication origin function is characterized by two parameters, the origin activation time and origin efficiency. Each replication origin initiates at a set, reproducible time in S phase, and replication origins are broadly classified as either early-activating or late-activating. Origin efficiency is a measure of how likely an origin is utilized in a given S phase. Replication origins can be active in every cell cycle, active only in a fraction of cell cycles, or not active during normal S phase. These inactive, dormant origins
are typically only utilized if the cell encounters replicative stress and needs initiation from additional replication origins to compensate for stalled replication forks (Blow et al. 2011).

Replication origin activation time can be measured by detecting the presence of newly replicated DNA, which can be distinguished from parental DNA by labeling the available nucleotide pools with a distinct isotope (Meselson and Stahl 1958; Raghuraman et al. 2001). Labeled nucleotides are incorporated into nascent strands, demarcating newly replicated genomic regions. At the beginning of S phase, only regions proximal to early-activating replication origins will harbor labeled, nascent DNA. As S phase progresses, this labeling will be extended to regions neighboring late-activating replication origins. Thus, the temporal appearance of labeled nucleotides reveals origin activation time.

The replication timing profile can also be uncovered by tracking DNA copy number changes throughout S phase (Yabuki et al. 2002; Müller et al. 2014). Once a genomic region is replicated, its relative copy number doubles. The increase in relative copy number within a population can be tracked by synchronously releasing the population from a G1 arrest and sampling the total DNA throughout S phase. Evaluating the temporal increase in DNA copy number at replication origins yields an estimate of origin activation time.

As obtaining a synchronous population of cells may be unsuitable, or if arresting the cell population might alter the replication time in S phase, replication timing profiles can be derived by assessing the DNA copy number in S-phase-sorted cells (Schübeler et al. 2002; Koren et al. 2010; Müller et al. 2014). S-phase cells can be recovered from an asynchronous population by fluorescently-activated cell sorting (FACS) using total DNA content as a marker. This enriched population contains an equal mixture of cells in early and late S phase, and thus the recovered DNA depicts the replication timing profile of an average cell in the middle of S phase. By
normalizing this data to the copy number present in G1 or G2, the replication time of each genomic region, and subsequently the replication origin activation time, can be inferred.

In contrast to revealing the entire replication timing profile, alternative techniques focus on mapping only early-activating replication origins. These experiments rely on the chemical compound hydroxyurea (HU), which depletes cellular nucleotide pools (Slater 1973), thus activating the intra-S–phase checkpoint and restricting replication to only regions proximal to early replication origins (Weinert et al. 1994; Santocanale and Diffley 1998; Lopes et al. 2001).

One method to recover these early-activating regions is to enrich for single-stranded DNA (ssDNA) (Feng et al. 2006). Replication fork progression requires the unwinding of DNA, and thus genomic regions actively undergoing replication contain ssDNA. These ssDNA regions can be selectively enriched by their enhanced affinity for random primers relative to double-stranded DNA. By extending these primers with a fluorescently-labeled nucleotide, ssDNA regions can be subsequently recovered. In the presence of HU, replication forks stall in close proximity to early-activating replication origins, causing an enrichment of ssDNA at these locations.

Instead of recovering ssDNA, newly replicated DNA can be selectively enriched by inclusion of a noncanonical nucleoside analog such as bromodeoxyuridine (BrdU) (Crabbé et al. 2010). BrdU mimics thymidine, and can be incorporated into the genome during replication. Whereas parental sequences will contain only canonical nucleotides, nascent sequences can be selectively recovered by utilizing an antibody specific to BrdU. When cells are exposed to BrdU in the presence of HU, BrdU incorporation is limited to the sequences proximal to early-activating replication origins. BrdU incorporation can be extended to all possible replication origins by utilizing a mutant deficient in the intra-S–phase checkpoint (Katou et al. 2003).

Although each of these methods provide insights into replication origin activation
time, they are not adequately suited to profile activation efficiency. Origin efficiency can be measured by determining how often a diverging replication fork proceeds from a replication origin. In the fraction of cells where a replication origin is active, replication forks proceed bidirectionally from the replication origin. This results in two opposite-moving replication forks diverging from the replication origin. In contrast, in the population subset where the replication origin is inactive, the sequences surrounding the replication origin are copied by a neighboring replication fork moving in only one direction. By comparing the frequency of these two occurrences, an origin efficiency metric can be assigned to each replication origin.

Replication fork movement can be inferred by mapping Okazaki fragments across the genome (Smith and Whitehouse 2012; McGuffee et al. 2013). Okazaki fragments are only present on the lagging strand, and thus their enrichment on a particular strand reveals the direction of replication fork movement. By utilizing a mutant deficient in ligating together neighboring Okazaki fragments, these intermediates can be selectively isolated by size and subsequently sequenced. Leftward-moving replication forks place Okazaki fragments on the Watson strand, and rightward-moving replication forks position Okazaki fragments on the Crick strand. Thus, at active, highly-efficient replication origins, diverging replication forks result in an abrupt strand switch of Okazaki fragments from the Watson strand to the Crick strand. In contrast, inefficient replication origins tend to have Okazaki fragments predominantly on the same strand surrounding the replication origin, revealing that replication forks travel in only one preferred direction through the replication origin. The Okazaki fragment distribution around most replication origins is a combination of both possible events, and deconvolving the signal yields intermediate efficiency values between 0 and 1.
1.3.6 Chromatin structure surrounding replication origins

The chromatin architecture at replication origins was initially described at individual replication initiation sites. A DNase digestion assay at ARS1 demonstrated that nucleosomes are well-positioned around the ACS (Thoma et al. 1984), and this chromatin configuration is primarily dependent on both ORC and Abf1 serving as barrier elements (Lipford and Bell 2001). Loss of ORC binding by mutating the ACS leads to nucleosome occlusion over the replication origin (Lipford and Bell 2001). Similarly, deletion of the Abf1-binding site shifts nucleosomes closer to the ACS (Hu et al. 1999; Venditti et al. 1994), impairing replication function (Marahrens and Stillman 1992). The barrier role of Abf1 can also be fulfilled by Rap1 or Gal4 without a loss of replication activity (Marahrens and Stillman 1992).

This nucleosome configuration around ARS1 is critical for replication function. Progressively moving the upstream nucleosome closer to the ACS hinders replication function, presumably by sterically interfering with ORC binding (Simpson 1990). Interestingly, artificially moving the upstream nucleosome further away from the ACS also impairs replication function (Lipford and Bell 2001). Even though this chromatin structure permits ORC binding, Mcm2-7 loading is significantly diminished; thus, nucleosome positioning plays a role in both ORC recruitment and pre-RC assembly.

More recent genome-wide nucleosome maps demonstrated that the chromatin configuration at ARS1 extends to replication origins genome-wide. Nucleosomes are depleted at replication origins, establishing a ~125-bp nucleosome-free region (Eaton et al. 2010; Berbenetz et al. 2010). Nucleosomes flanking the replication origin are asymmetrically positioned around the ACS, placing the upstream nucleosome in closer proximity to the ACS. Distal nucleosomes on both sides of the ACS are also consistently phased. There is some variability in the width of the nucleosome-free region which has been shown to affect activation time; narrow and wide NFRs tend
to be later-activating than NFRs closer to the canonical 125-bp width (Berbenetz et al. 2010).

Replication origins are kept nucleosome-free in part by sequence, as nucleosomes are excluded from replication origins when they are assembled in vitro (Kaplan et al. 2009; Eaton et al. 2010). In addition to the T-rich ACS sequence that discourages nucleosome formation, replication origins contain stretches of nucleosome-excluding A-rich sequences downstream of the ACS (Eaton et al. 2010). These sequences are likely an important complement to the ACS, as high-scoring ACS sites lacking these sequences do not serve as replication origins. The chromatin structure is vastly different at these non-functional ACS sites: nucleosomes are much more occluded over the replication origins and exhibit no consensus positioning around the ACS.

Although replication origins are inherently nucleosome-free, ORC binding is required to precisely position the flanking nucleosomes (Eaton et al. 2010; Berbenetz et al. 2010). The absence of ORC leads to nucleosome encroachment over the ACS and loss of nucleosome phasing at locations more distal from the ACS. This suggests that additional factors such as chromatin remodelers may play a role in establishing the chromatin environment around replication origins. In support of this view, the chromatin remodelers INO80 and ISW2 have been shown to localize to replication origins (Shimada et al. 2008; Vincent et al. 2008). Addition of CHRAC, a member of the ISWI chromatin remodeler family, enhanced replication initiation from the SV40 replication origin in vitro (Alexiadis et al. 1998). Knockdown of the SWI/SNF chromatin remodeler decreases the stability of a plasmid harboring ARS121, although plasmids containing the ARS1, ARS307, or ARS309 templates are unaffected, suggesting chromatin remodelers selectively target specific replication origins (Flanagan and Peterson 1999).
In addition to the role of chromatin structure in defining replication origins, the chromatin environment likely impacts replication origin activation. Replication origins located in the telomere regions are late-replicating (Raghuraman et al. 2001), and this is likely mediated by the compacted chromatin structure that is responsible for gene silencing. Artificial movement of ARS1, an early-replicating origin, into the telomere region delays its replication activation time to later in S phase; conversely, a late replication origin located in the telomere could activate earlier if removed from its endogenous chromosomal location (Ferguson and Fangman 1992). As replication initiation factors like Cdc45 are relatively limiting (Mantiero et al. 2011), replication activation time is impacted by competition for access to these proteins. Thus, a favorable chromatin environment can establish conditions for sequestering these replication initiation proteins.

Several experiments have suggested that histone acetylation positively influences origin activation. In yeast, histone acetylation is downregulated by the histone deacetylase Rpd3 (Vogelauer et al. 2002). In the absence of Rpd3, histone acetylation is subsequently increased, leading to the earlier activation of constitutively late replication origins (Vogelauer et al. 2002; Aparicio et al. 2004; Knott et al. 2009). Local assays have also shown the impact of histone acetylation on replication timing, as targeting a histone acetyltransferase to a late origin leads to an earlier S-phase activation time (Vogelauer et al. 2002). Around ARS1, not only are multiple sites of histone acetylation detected on H3 and H4, but this acetylation is specifically enhanced during S phase (Unnikrishnan et al. 2010). However, no correlation was found between genome-wide acetylation levels and replication timing (Nieduszynski et al. 2006), calling into question the direct role of histone acetylation in origin activation. An alternative model posits that the alternative chromatin compaction
role of Rpd3 at the repetitive rDNA locus regulates origin activation time genomewide (Smith et al. 1999; Yoshida et al. 2014). Rpd3 deletion leads to increased silencing of the rDNA locus and adversely impacts origin activation from rDNA replication origins (Yoshida et al. 2014). Since chromosomal and rDNA replication origins compete for limiting replication initiation factors (Kwan et al. 2013), silencing at the rDNA locus via Rpd3 deletion may restrict the sequestration of these factors to rDNA replication origins. This would enable enhanced recruitment of replication initiation factors to chromosomal replication origins, promoting their earlier activation through an acetylation-independent mechanism.

As the yeast genome is relatively compact, the intergenic regions where replication origins reside are typically in close proximity to gene bodies. Thus, regulation of the transcription program can have a direct impact on replication function. Replication origins near TSSs tend to be earlier firing; conversely, replication origins near gene ends are later firing (Berbenetz et al. 2010). The earlier firing near TSSs may be due to the permissive chromatin environment established for active transcription, or by the local proximity of transcription factors that promote origin activation. As one example, Fkh1/2 binding has been identified as promoting earlier origin activation by both spatially clustering replication origins in close proximity within the nucleus and directing Cdc45 localization to these areas (Knott et al. 2012). This spatial clustering has also been observed with replication origins near centromeres, which replicate early in S phase (Pohl et al. 2012).
2

Epigenome characterization at single base-pair resolution

2.1 Introduction

Short-read deep-sequencing technologies have the potential to revolutionize epigenomic profiling by making it possible to map DNA fragments with single base-pair resolution at reasonable cost. This ideal has been achieved for nucleosomes, which can be mapped at high resolution by treatment with micrococcal nuclease (MNase). MNase is a single-strand–specific secreted glycoprotein that cleaves one strand when DNA breathes, then cleaves the other strand, resulting in a double-strand break. MNase then “nibbles” on the exposed DNA ends until it encounters an obstruction, such as a nucleosome, where the histone cores protect the DNA from further encroachment. Although MNase has long been used for studying nucleosomes (Noll 1974), its mechanism of action on DNA suggests that it will stop “nibbling” at any

Chapter 2 was modified from the following publication in Proceedings of the National Academy of Sciences 108(45): 18318–18323 (2011), authored by Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, and Henikoff S.
obstruction, such as a DNA-binding protein. As an example, MNase digestion of *Drosophila* nuclei followed by low-salt native chromatin extraction can be used to map both nucleosomes and paused RNA Polymerase II using paired-end sequencing (Weber *et al.* 2010). Similarly, Kent *et al.* (2011) used MNase digestion of uncrosslinked yeast nuclei to map binding sites for both nucleosomes and smaller particles identified as sequence-specific transcription factors. These studies showed that MNase-protected DNA fragments as small as ~50 bp could be recovered and mapped.

A limitation of using paired-end sequencing as a read-out method for MNase mapping is that standard sequencing library preparation methods are optimized for DNA fragments of a few hundred base pairs, whereas the fragments protected by DNA-binding proteins are an order-of-magnitude smaller. Here we introduce a rapid Illumina library preparation protocol that efficiently recovers particles down to ~25 bp in size but excludes primers and adapters without gel purification. We have applied this protocol to small amounts of DNA extracted from MNase-treated yeast nuclei and native chromatin, and have obtained occupancy maps that resolve nucleosomes and small subnucleosomal particles at or near single base-pair resolution. We show that two-dimensional plots of fragment coordinates vs. length can reveal chromatin features in the vicinity of transcription-factor–binding sites.

2.2 Results

2.2.1 Chromatin occupancy profiling via micrococcal nuclease (MNase) digestion

Genome-wide chromatin profiling with MNase has typically focused exclusively on nucleosome-sized DNA fragments. After permeabilization of the nuclear membrane, chromatin is subjected to an MNase digestion, resulting in the conversion of a contiguous chromosome into smaller DNA components. The MNase digestion products are typically separated and visualized via gel electrophoresis, where the major di-
gestion products around 150 bp, 300 bp, and 450 bp reveal the primary existence of mononucleosome, dinucleosome, and trinucleosome species, respectively (Noll 1974; Noll et al. 1975). However, typically only the genomic sequences protected by individual nucleosomes are selectively sequenced by gel excising the mononucleosome band at \( \sim 150 \) bp (Yuan et al. 2005; Albert et al. 2007).

Although this approach successfully reveals nucleosome positioning throughout the genome, the sample preparation restricts chromatin profiling to only one size range. Several groups reasoned that omitting the gel extraction step and isolation of only the \( \sim 150 \)-bp mononucleosome layer would result in a more comprehensive chromatin profiling across the genome (Floer et al. 2010; Weber et al. 2010; Kent et al. 2011). In addition to recovering higher molecular weight species indicative of dinucleosomes and trinucleosomes (Kent et al. 2011), smaller “subnucleosome” fragments (<140 bp) are now evident across the genome (Weber et al. 2010). These species are enriched at partially unwrapped nucleosome locations (Floer et al. 2010) and transcription-factor–binding sites (Kent et al. 2011).

We reasoned that by simultaneously resolving both nucleosome and “subnucleosome” fragments, this extension of the classical MNase digestion assay could effectively probe the genome-wide chromatin structure in only one experiment. In conjunction with previously annotated DNA-binding motifs and gene annotations, one MNase experiment would reveal the same chromatin occupancy information previously only obtained by combining MNase, DNase, and many ChIP-seq datasets. Further, the isolation of smaller fragment sizes, particularly in the range of 25–120 bp, could significantly improve the spatial resolution of transcription-factor binding over chromatin immunoprecipitation.

As budding yeast is a relatively compact and well-annotated genome (Goffeau et al. 1996; Harbison et al. 2004; MacIsaac et al. 2006), we generated chromatin occupancy profiles from an asynchronous \( S. \) cerevisiae population in log-phase growth.
Figure 2.1: MNase primarily digests the linker DNA regions between nucleosomes. Chromatin from asynchronous, log-phase growth *S. cerevisiae* was digested with MNase for 2.5 (light) and 20 (complete) minutes. The size of distinct DNA bands corresponds to the nucleosome array length.

We then subjected the native, un-crosslinked chromatin to two MNase digestion conditions, a 2.5-minute “light” and a 20-minute “complete” digestion (Fig. 2.1). Previous groups demonstrated that nucleosomes differ in their ability to protect against an MNase digestion, which can serve as a proxy for their inherent stability in the genome (Bryant *et al.* 2008; Weiner *et al.* 2010; Xi *et al.* 2011). We therefore hypothesized that performing two digestion conditions on the entire chromatin profile could extend these protein dynamic analyses to transcription factors as well.

After isolating the protected DNA fragments, we then prepared libraries for paired-end sequencing on the Illumina platform. The Illumina platform resulted in a DNA sequence read length of 25 bp from either end of the DNA fragments, which were then subsequently aligned to the yeast genome with Bowtie (Langmead *et al.* 2009). Although the Illumina sequencing platform is designed for 150-bp fragments (Fan *et al.* 2010), we recovered a range of fragments from 25–400 bp within each MNase digestion condition (Fig. 2.2); in particular, a significant fraction (>30%)
Figure 2.2: MNase fragment length profile. A range of DNA fragments, primarily 25–400 bp, are recovered from the Illumina sequencing platform for both light (2.5') and complete (20') MNase digestions. Fragments are classified as Subnucleosome (25–120 bp), Nucleosome (150–175 bp), or Dinucleosome (300–360 bp).

of the reads were smaller than 120 bp, indicating successful capture of subnucleosome fragments. Both digestion conditions yielded recovery of primarily mononucleosome-sized fragments with peaks of 165 and 153 bp for the light and complete digestions, respectively. Minor peaks at 325 and 303 bp reflect retrieval of fragments occupied by two sequential nucleosomes.

2.2.2 Enhanced visualization of the chromatin landscape

Genomics data is typically visualized by plotting the one-dimensional density of aligned DNA fragments along each chromosome, indicating locus-specific regions of enriched signal (Pepke et al. 2009). However, in addition to the spatial location
Figure 2.3: Two-dimensional visualization schematic of paired-end MNase-digested DNA fragments. The x-axis corresponds to the chromosomal coordinates of the DNA fragment, and the y-axis reflects the fragment length.

along the genome, the inferred fragment widths from our paired-end sequencing data yield additional insights into the type of protein occupying the protected region. To preserve this information, we utilized a two-dimensional plotting scheme (Floer et al. 2010) where each recovered DNA fragment is depicted as a rectangle (Fig. 2.3). The rectangle spans the genomic region between the start and end alignment positions of the DNA fragment, and the DNA fragment width determines the y-axis position of the rectangle. We fill each rectangle with a light-blue transparent color, so that
Figure 2.4: Visualization of the chromatin profile captures nucleosomes and small DNA-binding proteins. The genomic location of recovered DNA fragments are displayed as a function of their width for both light (2.5', top panel) and complete (20', middle panel) MNase digestions. Clusters of fragments in the 140–175 bp denote nucleosomes, whereas subnucleosome fragment groups at 25–120 bp likely reflect protection of transcription factors. (Bottom panel) The fragment length profiles at the inferred nucleosome positions are displayed for both light (2.5', red) and complete (20', green) MNase digestions. The -1 nucleosome directly upstream of YCR050C is particularly sensitive to MNase digestion.

Clusters of DNA fragments with coinciding coordinates appear as darker blue in the resulting plot. As longer DNA fragments tend to overlap each other, we improved the resolution of distinct read clusters by halving the span of each rectangle.

We visualized the aligned DNA fragments across the genome from both the light (2.5') and complete (20') MNase digestions at a representative genomic lo-
Individual nucleosomes across the genome typically contain a range of DNA fragments centered around ~165 bp for the “light” digestion and ~150 bp for the “complete” digestion. However, the nucleosome centered at position 213,830 shows a different fragment length distribution profile (Fig. 2.4, bottom panel, highlighted). While the 2.5’ MNase digestion yielded a fragment length profile similar to other nucleosomes, the fragment length distribution at this nucleosome from the 20’ MNase digestion is more variable and contains markedly smaller fragments. This differential sensitivity to MNase digestions is likely an example of a “fragile” nucleosome (Weiner et al. 2010; Xi et al. 2011). These nucleosomes are less resistant to prolonged MNase digestion and constitute ~5% of nucleosomes genome-wide (Xi et al. 2011). Genes with high expression rates tend to contain these sensitive nucleosomes in their promoter regions, possibly reflecting high nucleosome turnover rates associated with constitutive RNA Pol II occupancy.

2.2.3 Chromatin footprinting at the UASg reveals a non-canonical nucleosome protection

One of the most highly studied promoter regions in S. cerevisiae is the UASg, situated between the GAL10 and GAL1 genes. These genes are silenced during normal
growth conditions on a sugar source such as raffinose or glucose, with well-positioned, phased nucleosomes occupying the promoter region (Lohr 1984; Bryant et al. 2008). However, in the presence of galactose, nucleosomes are evicted and these genes are rapidly upregulated (Douglas and Hawthorne 1966). The transcriptional activator Gal4 recognizes four binding sites in the UASg and plays a crucial role in remodeling the silenced chromatin architecture into a more permissive state for transcriptional activation (Guarente et al. 1982). In the absence of galactose, these binding sites are primarily obscured by a nucleosome located precisely over the UASg (Lohr 1984; Bryant et al. 2008). MNase digestion at this particular nucleosome yields a smaller footprint than typical nucleosomes, suggesting that this nucleosome is partially unwrapped. A series of ChIP experiments coupled with mutational analyses conclusively showed that the RSC chromatin remodeler is highly localized to this nucleosome (Floer et al. 2010). RSC interaction with this nucleosome likely alters the canonical 147-bp protection typically conferred by nucleosomes, ensuring that the Gal4 binding sites remain partially accessible even when galactose is not present.

We utilized our chromatin landscape maps to yield additional insights into how the chromatin state is primed for upregulation (Fig. 2.5). At the UASg, we confirmed the smaller nucleosome protection in both digestion conditions, with a median fragment length of 132 bp and 113 bp for the 2.5′ and 20′ MNase digestions, respectively (Fig. 2.5, bottom panel). Even though the majority of the fragments in each digestion condition are above 100 bp at this location, the two-dimensional fragment plots reveal two smaller, yet distinct, digestion products. This suggests occupancy of small DNA-binding factors directly adjacent to the partially-unwrapped nucleosome. Extending the fragment boundaries from each cluster toward the x-axis reveals that these fragments originated at either the left or right nucleosome boundary, respectively. This agrees with previous footprinting experiments, which identified the ~18-bp linker on either side of the UASg nucleosome as hypersensitive to nuclease
Figure 2.5: RSC binding is evident as a partially-unwrapped nucleosome at the UASg. The genomic location of recovered DNA fragments are displayed as a function of their width for both light (2.5’, top panel) and complete (20’, middle panel) MNase digestions. Clusters of fragments in the 140–175 bp range denote nucleosomes. (Bottom panel) The fragment length profiles at the inferred nucleosome positions are displayed for both light (2.5’, red) and complete (20’, green) MNase digestions. The nucleosome positioned over the UASg has a smaller fragment distribution than typical nucleosomes. Filled rectangles in the gene schematic denote the UASg (black), along with the RSC (purple) and Gal4 (green) binding sites.

digestion (Fedor et al. 1988; Bryant et al. 2008; Floer et al. 2010). Interestingly, the left UASg sequence (more proximal to GAL10) contains three Gal4 binding sites, and the right UASg sequence (more proximal to GAL1) harbors three RSC binding sites (Floer et al. 2010). Thus, the size ranges and genomic positions of recovered
fragments likely reflect occupancy of three distinct proteins: (1) ~120-bp fragments correspond to protection by a partially-unwrapped nucleosome; (2) smaller fragments at the UASg proximal to GAL10 reflect Gal4 occupancy; and (3) smaller fragments at the UASg proximal to GAL1 signify RSC binding.

2.2.4 **Proximal gene promoters are consistently occupied by transcription factors**

The nucleosome structure around gene promoters has been studied extensively, revealing that most gene promoters contain a ~140-bp nucleosome-free region (NFR) upstream of the TSS (Lee *et al.* 2004; Sekinger *et al.* 2005; Yuan *et al.* 2005; Shivashwamy *et al.* 2008; Mavrich *et al.* 2008a). This region is highly-enriched for many transcription-factor–binding motifs (Harbison *et al.* 2004), and there is a high concentration of DNase hypersensitive sites within this regulatory region (Hesselberth *et al.* 2009). Previous MNase chromatin profiling experiments showed small fragment enrichment within this nucleosome-free region at gene promoters in aggregate (Weber *et al.* 2010; Kent *et al.* 2011), but did not evaluate the conservation of this pattern across individual genes.

We extended our analysis of the chromatin structure at individual loci to all promoter regions with a well-annotated transcription start site (Xu *et al.* 2009). We then obtained chromatin occupancy profiles from the 2.5′ MNase digestion across the ~5,000 promoter regions centered on the TSS and oriented by the direction of transcription. By aggregating all the individual chromatin landscapes, we first generated a consensus view of the yeast promoter chromatin architecture (Fig. 2.6). As before, individual clusters at 140–175 bp reflect protection by nucleosomes. In line with previous estimates (Mavrich *et al.* 2008a), the consensus -1 and +1 nucleosomes are positioned at -220 bp and 45 bp relative to the TSS, respectively. In contrast, there is a diffuse signal of smaller fragments concentrated in the ~120-bp NFR directly upstream of the TSS. Interestingly, there is a correlation between the width of
Figure 2.6: Average promoter chromatin architecture shows DNA-binding proteins occupy the NFR. The genomic location of recovered DNA fragments are displayed as a function of their width for the light (2.5', top) MNase digestions across 4,958 yeast TSS (Xu et al. 2009). Clusters of fragments in the 140–175 bp range denote nucleosomes, whereas smaller 25- to 120-bp fragments likely reflect protection of transcription factors in the promoter NFR.

The smaller fragments and their genomic position relative to the TSS; even though the ~50-bp fragments occupy the center of the NFR, the ~120-bp fragments are located closer to the upstream nucleosome. We also detected super-nucleosomal reads (175–200 bp) above both nucleosomes flanking the TSS, likely reflecting recovery of a longer DNA fragment spanning the NFR and either nucleosome.

To facilitate comparison between individual gene promoters, we first subsetted the fragments into separate nucleosome (140–175 bp) and subnucleosome (25–120 bp) classes. Within each group, we extracted the fragment midpoint positions and con-
Figure 2.7: Transcription level does not dramatically impact promoter chromatin architecture. The subnucleosome (25- to 120-bp fragments, left) and nucleosome (140- to 175-bp fragments, right) tracks from the light (2.5' MNase digestion, ordered by transcription level, are displayed for each of the 4,958 yeast TSS (Xu et al. 2009) (bottom). Average chromatin profiles are obtained by grouping TSS by transcription level (top).

constructed a density occupancy profile across each promoter. Viewing all of these promoter profiles together in a heatmap reveals a consistent small protein occupancy confined to a narrow window around -100 bp relative to the TSS (Fig. 2.7, left). These footprints likely represent transcription factors (Harbison et al. 2004), as an MNase-resistant footprint was not previously recovered for the RNA polymerase complex (Bryant et al. 2008). Smaller fragments could also be found further upstream; however, their positioning is not well-conserved between promoters. Promoter oc-
cupancy levels are slightly dependent on transcription rate; however, a substantial small fragment occupancy is present even at low-expressing and silent genes. In contrast, the gene body is almost entirely devoid of small fragment footprints.

The nucleosome occupancy profiles show canonical dyad positioning of the +1 nucleosome just downstream of the TSS, and consistent locations of the +2 and +3 nucleosomes at ~200 bp and ~400 bp, respectively (Fig. 2.7, right). The nucleosomes within the gene body exhibit slightly less occupancy for the highest-expressing genes. Even though the NFR upstream of the TSS is well-defined across all promoters, the -1 and -2 nucleosome positions are more variable. In particular, higher-expressing genes position these nucleosomes slightly further from the TSS relative to low-expressing and silent genes. This additional flexibility in the upstream chromatin architecture may reflect gene-specific TF-binding in the promoter region, revealing additional modes of transcriptional regulation.

2.2.5 Transcription-factor binding directs nucleosome positioning

Even though most gene promoters harbor protein footprints, individual transcription factors are typically confined to a much smaller gene subset by specific regulatory sequences in the proximal promoter region (Harbison et al. 2004; MacIsaac et al. 2006). We utilized our chromatin occupancy profiles to ask if there are different chromatin architecture patterns associated with individual transcription factors, and how effective these transcription factors protected their regulatory sequences when exposed to the extended MNase digestion time (20').

Abf1

As a proof of principle, we focused on the transcription factor Abf1p, an essential protein involved in DNA replication and transcription regulation (Rhode et al. 1989; Lee et al. 2002; Miyake et al. 2004). Abf1 recognizes a 13-bp sequence (Della Seta
et al. 1990; Gordân et al. 2011), present at over 12,000 genomic locations (De Boer and Hughes 2012), yet Abf1-binding is only detected at several hundred sites in vivo (Harbison et al. 2004; MacIsaac et al. 2006). To further refine the Abf1-binding sites at high resolution, we first performed standard chromatin immunoprecipitation coupled to Illumina sequencing (ChIP-seq) of a Myc-tagged Abf1. We called standard ChIP-seq peaks using MACS (Zhang et al. 2008), and then scanned each ChIP-enriched region searching for a high-quality match to the Abf1 motif (MacIsaac et al. 2006). This yielded a high-confidence set of 409 Abf1-bound locations.

We generated chromatin occupancy profiles for the “light” and “complete” MNase digestions at each individual Abf1 location. We then aggregated the occupancy profiles together across all Abf1 sites, yielding a consensus view of the chromatin architecture from each MNase digestion condition (Fig. 2.8). Individual nucleosomes are evident as distinct clusters from 140–175 bp in both digestion conditions, centered at ±140 bp and ±320 bp relative to the Abf1 motif center. This nucleosome spacing creates a 130-bp nucleosome-free region, 10 bp longer than the typical promoter NFR. The linker region between the distal and proximal nucleosomes spans 30 bp, also wider than the canonical 23-bp distance between nucleosomes over gene bodies.

Whereas nucleosome protection results in confined spherical clusters of protected fragments, the protection pattern over the Abf1 motif resembles an inverted triangle (Fig. 2.8D). The lower vertex of the triangle begins at 25 bp, the smallest fragments we recover, and represents precise MNase cleavage on each side of the bound Abf1 protein. Less digested fragments have a longer fragment length, yet the fragments always span the motif position and remain concentrated between two white diagonal lines. These white diagonal lines represent irrecoverable fragments because of protection by the Abf1 footprint, as any fragments in this region would require MNase cleavage within the Abf1-binding site. Thus, MNase digestion is restricted to the linker region between Abf1 and the flanking nucleosomes where no DNA-
Figure 2.8: A high-resolution footprint is revealed at Abf1-binding sites. The genomic location of recovered DNA fragments are displayed as a function of their width for the light (2.5', A) and complete (20', B) MNase digestions across 409 Abf1 motif locations. (C) Chromatin profiles from the 2.5' MNase digestion time (green) and 20' MNase digestion time (red) are merged into one plot. Clusters of fragments in the 140- to 175-bp range denote nucleosomes, whereas smaller 25- to 120-bp fragments likely reflect protection of Abf1. (D) Schematic model of DNA fragments at the Abf1-binding site. MNase cleavage is restricted to the linker region between Abf1 and the flanking nucleosomes. DNA fragments that span the Abf1 footprint (blue rectangles) fill an inverted triangular region (between green dotted lines). Because the Abf1 footprint prevents MNase cleavage within its binding site, a subset of reads are absent (white rectangle).
binding proteins reside. In the 20′ MNase digestion, there is a high concentration of fragments along the diagonal boundary. Because of the prolonged digestion time, MNase has extensively cleaved all the linker DNA between Abf1 and one flanking nucleosome, resulting in one edge of the recovered fragments aligning directly with the Abf1 footprint boundary. Most (~70%) of the fragments spanning the motif are <130 bp, further confirming the NFR width estimation.

The two MNase digestion conditions also result in different types of fragments spanning the adjacent nucleosomes. Similar to the aggregate promoter chromatin profiles, larger supernucleosomal (175–200 bp) fragments are only present in the 2.5′ MNase digestion. These fragments likely represent a recovered DNA fragment protected by both Abf1 and one flanking nucleosome, as the short digestion time prevented MNase cleavage in the linker region between the proteins. In the 20′ MNase digestion, 90–140 bp fragments are present directly below the nucleosomes. These likely reflect preferential MNase cleavage positions within the nucleosomal core, which can occur at a higher rate during a prolonged MNase digestion (Cockell et al. 1983). These cleavage products exist at similar proportions across the four nucleosomes flanking the Abf1 motif.

We next evaluated the Abf1 occupancy and surrounding nucleosome structure for individual motif locations from the 2.5′ MNase digestion condition. As before, we collapsed all the small (25- to 120-bp) and nucleosome-sized (140- to 175-bp) fragments into respective one-dimensional signal tracks (Fig. 2.9). Well-positioned nucleosomes flanking the Abf1 binding site are evident. Consistent with the barrier model of nucleosome positioning (Kornberg 1981; Kornberg and Stryer 1988), more variability is present in the ±2 nucleosome positioning as compared to the ±1 nucleosome. Profiling the small MNase-resistant fragments reveals a footprint concentrated over the Abf1 motif, resolving the binding site to a higher resolution relative to ChIP-seq. Even though additional small fragment footprints decorate the local chromatin
Figure 2.9: Abf1 occupancy positions adjacent nucleosomes. The nucleosome (140- to 175-bp fragments, left) and subnucleosome (25- to 120-bp fragments, right) tracks from the light (2.5') MNase digestion are displayed for each of 409 Abf1-binding sites. The Abf1 footprint is better resolved by the subnucleosome signal (black) as compared to ChIP-seq (green).

environment surrounding Abf1-binding sites, only a single protein footprint primarily resides within the NFR harboring the Abf1 motif. Abf1 occupancy thus establishes a stereotypical chromatin configuration that positions the flanking nucleosomes.

Rap1

We next investigated the chromatin architecture surrounding Rap1-binding sites. Rap1 is an essential transcription factor involved in myriad genomic functions including telomere maintenance, transcription at the mating type loci, and ribosomal
Figure 2.10: Two distinct chromatin configurations impact Rap1 residency. The nucleosome (140- to 175-bp fragments, left) and subnucleosome (25- to 120-bp fragments, right) tracks from the light (2.5') MNase digestion, ordered by Rap1 residence time (Lickwar et al. 2012), are displayed for each of the 278 Rap1 sites. Average chromatin profiles are separated into two groups, low residence (<72 min, red) and high residence (>72 min, green).

gene regulation (Lieb et al. 2001; Piña et al. 2003). The genomic binding locations of Rap1 have recently been interrogated by high-resolution ChIP-exo (Rhee and Pugh 2011), and the chromatin state at these binding sites substantially influences Rap1 turnover (Lickwar et al. 2012). We asked if our chromatin occupancy profiles could yield additional insights into the factors that affect Rap1 binding.

We utilized the Rap1 position-weight matrix (PWM) (MacIsaac et al. 2006) to identify Rap1-binding sites within high-confidence ChIP-exo peaks (Rhee and Pugh
After overlaying the Rap1-binding kinetic data derived from a competition assay (Lickwar et al. 2012), we recovered 278 distinct Rap1-bound sequences. As with Abf1, we consolidated the nucleosome and small fragment footprint signals into individual one-dimensional signal tracks, respectively, across each Rap1-binding site. Rap1-binding sites were then ordered by Rap1 residency time to evaluate the relationship between turnover dynamics and MNase protection, orienting by the strand containing the CA-rich motif (Fig. 2.10). In the 2.5' MNase digestion profile, a small fragment footprint is clearly evident at the motif position. Surprisingly, ordering by the residence time reveals two types of chromatin architectures downstream of the Rap1-binding site. Rap1-binding sites with a long residence time (>72 minutes) harbor a downstream, secondary footprint, which is noticeably absent from Rap1 sites with more rapid turnover. These stable Rap1-binding sites are proximal to highly-transcribed ribosomal genes, possibly because prolonged Rap1 residency increases recruitment of the transcriptional machinery (Lickwar et al. 2012).

To more extensively compare the chromatin architecture between the two residency clusters, we generated separate aggregate two-dimensional chromatin occupancy profiles across all the binding sites within each turnover class (Fig. 2.11). The Rap1-binding sites with shorter residency display a similar MNase digestion profile to the Abf1-binding sites. Each flanking nucleosome is located at 135 bp from the motif position, establishing an NFR spanning 120 bp. This NFR width is further confirmed by the fragment length distribution of recovered reads at the motif position, which are mostly confined to <120 bp. The consensus distance of the ±2 nucleosome from the motif position is 315 bp, yielding a 30-bp linker distance between the proximal and distal nucleosomes as was observed for Abf1.

In contrast, Rap1-binding sites with a long residency time yielded a much different chromatin occupancy profile. The nucleosome-free region is significantly wider at 220 bp. In addition to the small fragment footprints spanning the Rap1 motif, there is a
Figure 2.11: Cofactor occupancy coincides with longer Rap1 residency. The genomic location of recovered DNA fragments are displayed as a function of their width for the low residence (<72 minutes, left) and high residence (>72 minutes, right) Rap1 sites from the 2.5’ MNase digestion.

substantial small fragment footprint 100 bp downstream of the Rap1 motif. Longer fragments (130–200 bp) encompassing both footprints provides further support for a distinct factor occupying the downstream region. This factor is relatively close to the downstream nucleosome, as a high concentration of supernucleosomal fragments are evident above this nucleosome. Interestingly, the linker region between the +1 and +2 nucleosomes is only 16 bp, much less than the 30-bp linker region between the two upstream nucleosomes. Thus, binding by both Rap1 and the downstream factor may limit the available space between the downstream nucleosomes.

2.3 Discussion

We have demonstrated that chromatin profiling via MNase digestion is greatly enhanced by eliminating the nucleosome size-selection step, leading to the recovery and subsequent sequencing of many fragment lengths. Short, MNase-resistant fragments
reveal the locations of smaller proteins, such as transcription factors, and provide high-resolution information on the surrounding chromatin structure. Longer, supernucleosomal reads provide evidence of transcription-factor–binding proximal to a nucleosome, and may even yield evidence of a direct histone-TF interaction. Since no chromatin immunoprecipitation selection step is involved, MNase profiling yields an agnostic view of the entire genomic chromatin structure at a high resolution. Combined with sequence annotations and existing genomics data, it can reduce the need for hundreds of ChIP sample preparations to comprehensively evaluate the chromatin state.

2.3.1 Evaluating protein-binding dynamics by varying MNase digestion times

Classical footprinting experiments showed that resistance to nuclease cleavage serves as a proxy for protein occupancy (Ganter et al. 1993; Teng et al. 2001). By comparing protein footprints from short (2.5') and long (20') MNase digestion experiments, we could also gain further insights into protein-binding properties at locations throughout the genome. We confirmed the presence of a “fragile” nucleosome upstream of YCR050C (Fig. 2.4) and a constitutively unwrapped nucleosome at the UASg (Fig. 2.5). Several factors could affect this sensitivity to MNase digestion, including sequence-dependent binding interactions, post-translational modifications of the nucleosome, incorporation of histone variants, or recruitment of chromatin remodelers. These different modes of protein-binding likely play an important role in priming nearby genes for upregulation.

2.3.2 Chromatin architecture at gene promoters

MNase profiling of the nucleosome structure at gene promoters has shown a well-defined nucleosome-free region upstream of the transcription start site (Lee et al. 2004; Sekinger et al. 2005; Yuan et al. 2005; Shivaswamy et al. 2008; Mavrich et al. 2008).
2008a), and we confirmed this nucleosome depletion in the promoter region. Increased gene expression was correlated with a slightly wider NFR; although all genes positioned the dyad of their +1 nucleosome 50 bp downstream of the TSS, the consensus -1 nucleosome location was slightly more variable (Fig. 2.7). Silent and low-expressing genes contained an upstream nucleosome at -215 bp, medium-expressing genes at -223 bp, and highly-expressed genes at -230 bp. Interestingly, the nucleosome-depleted promoter region of silent genes contained a considerable small fragment footprint, indicating either the constitutive occupancy of a repressor, or a bound factor priming the gene for activation by maintaining a relatively open chromatin structure.

2.3.3 Small MNase-protected fragments reveal TF-binding footprints

Advances in sequencing technology have greatly increased the throughput of DNA footprinting experiments (Hesselberth et al. 2009; Kent et al. 2011). Our two-dimensional chromatin occupancy profiles yield further insights into transcription-factor binding by simultaneously profiling both nucleosomes and small DNA-binding factors. We showed that Abf1-binding sites contain a 130-bp nucleosome-free region, derived from both the flanking nucleosome positions and concentration of small fragments at the Abf1-motif position. The width of this nucleosome-free region is larger than the 120-bp consensus NFR width found at promoters, and the 30-bp linker region between the proximal ±1 and distal ±2 nucleosomes is almost twice as wide as the typical 18-bp yeast linker width found throughout the genome (Jansen and Verstrepen 2011). This nucleosome packing configuration surrounding a well-defined transcription-factor–binding site supports the barrier model of nucleosome positioning, which reasons that competition among all DNA-binding factors is a determinant of nucleosome spacing (Kornberg 1981; Mavrich et al. 2008a).

The MNase digestion profiles around Abf1-binding sites revealed several intrigu-
ing properties of the chromatin structure. First, small fragments spanning the Abf1-motif position are concentrated within a triangular region centered at the binding site (Fig. 2.8). As small fragments that do not span the Abf1-binding site are not recovered, this suggests a model where Abf1 typically binds the DNA independently. Although the inferred NFR is 130 bp, our MNase mapping shows a substantially smaller footprint, suggesting that the flanking nucleosomes are positioned by more than just steric hindrance with Abf1. Many Abf1-binding sites are located in close proximity to binding sites for the RSC chromatin remodeler, which may play a substantial role in actively positioning these nucleosomes (Badis et al. 2008).

Larger DNA fragments at 150 bp form distinct clusters at the inferred nucleosome positions; however, each nucleosome also contains a minor distribution of smaller 90- to 130-bp fragments. Even though DNA wrapping around nucleosomes provides a substantial barrier to nuclease cleavage, several MNase and DNase hypersensitive sites on DNA wrapped around the nucleosome core structure have been identified (Cockell et al. 1983; Tolstorukov et al. 2009; Winter et al. 2013). Nucleosomes containing a higher proportion of these auxiliary digestion products likely reflect partially-unwound or less stable nucleosomes (Weiner et al. 2010; Xi et al. 2011).

Most Rap1-binding sites displayed a similar chromatin architecture to Abf1; however, binding sites with a long Rap1 residence time (>72 minutes) yield a substantially different chromatin profile (Fig. 2.11). Unlike binding sites experiencing more frequent Rap1 turnover, genomic locations with an extended Rap1 residency contain an unusually wide NFR and a distinct small fragment footprint downstream of the Rap1-binding site. As a prolonged Rap1 presence leads to additional recruitment of the transcriptional machinery (Layer et al. 2010), this downstream factor may either directly interact with Rap1, enhancing its residence, or indirectly create a more favorable chromatin environment for Rap1 binding by displacing the +1 nucleosome.
2.4 Methods

2.4.1 Data deposition

The sequences reported in this chapter have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo, under accession number GSE30551.

2.4.2 Chromatin isolation

*S. cerevisiae* cultures were grown in YEPD medium at 30°C to OD\textsubscript{600} = 0.8. Nuclei were prepared as previously described (Furuyama and Biggins 2007), flash-frozen in liquid nitrogen, and stored at −80°C. Nuclei were thawed at room temperature and digested with MNase, followed by chromatin preparation as previously described (Furuyama and Biggins 2007), except that after MNase digestion, the slurry was passed four times through a 20-gauge needle, then four times through a 26-gauge needle (Jin and Felsenfeld 2007), and the combined S1+S2 supernatants were clarified by centrifugation in a fixed-angle Sorvall SS34 rotator at 17,200 × g at 4°C for 10 min at least twice or until no visible pellet remained.

2.4.3 Illumina sequencing and analysis

The standard Illumina paired-end library preparation protocol (Illumina # PE-930-1001) was used except as follows: (i) All Qiagen clean-up steps and gel purifications were omitted. (ii) Phenol/chloroform extraction was used to stop reactions followed by S300 spin column clean-up and Speed-vac volume reduction. (iii) Ampure XP bead steps following adapter ligation and PCR removed excess adapters and primers, respectively. (iv) Reaction volumes were 50µL, except for the PCR amplification step, where the volumes were 10 to 25µL. (v) PCR extensions were done at 60°C
to minimize bias against AT-rich sequences (López-Barragán et al. 2011). (vi) Low-retention 1.5 mL (siliconized) microfuge tubes were used throughout, minimizing tube transfers to reduce losses. A detailed protocol has been used for 20 to 800 ng starting DNA from MNase-treated chromatin (based on Quant-it Picogreen fluorescence with RNase added). Cluster generation, followed by 25 rounds of paired-end sequencing on an Illumina HiSeq 2000, was performed by the Fred Hutchinson Cancer Research Center Genomics Shared Resource. After processing and base calling by the Illumina Eland program, reads with zero, one, or two mismatches were mapped to the yeast genome using Bowtie (Langmead et al. 2009), where each multiple hit was assigned to one site chosen at random.

### 2.4.4 Construction of individual data signal tracks

To facilitate more consistent analysis between replicates and across conditions, aligned MNase reads were subdivided into two classes: nucleosome reads (fragment width of 140–175 bp) and small fragment reads (fragment width of 25–120 bp). Within each class, the midpoint position of every sequencing read was extracted, and a 20 bp bandwidth Gaussian kernel was constructed around each position to smooth the signal tracks. Then, individual density kernels were summed together to create an aggregate signal track across each chromosome.

### 2.4.5 Individual gene locus plots

Two-dimensional plots were constructed with the x-axis representing genomic position and the y-axis corresponding to the fragment length of the sequencing read. Each sequencing read is individually plotted by first extracting the start (st) and end (en) coordinates along with the fragment length (l). The sequence read is then trimmed to half its fragment length, reassigning the start and end positions to \( \text{start} = st + \frac{l}{4} \) and \( \text{end} = en - \frac{l}{4} \), respectively. A blue rectangle is then filled
with the following coordinates: \((\text{start}, l - \frac{1}{2}), (\text{start}, l + \frac{1}{2}), (\text{end}, l - \frac{1}{2}), (\text{end}, l + \frac{1}{2})\).

Reads are individually plotted so that denser blue regions correspond to the presence of more recovered reads. Aggregate two-dimensional plots (such as those shown in Fig. 2.6) are obtained by totaling the number of reads surrounding the feature position across multiple sites, with each aligned read oriented relative to the feature strand. Merge plots were constructed by assigning an RGB value to every square of the heatmap, where red represented the total signal from dataset one and green the signal from dataset two.

Cartoon schematics above each plot were computationally derived by finding peak positions in either the nucleosome (red) or small footprint (dark green) signal track. Nucleosome shading indicated the observed protein occupancy at a particular position, with darker red corresponding to a higher nucleosome read (150–175 bp) signal.
Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly

3.1 Introduction

Eukaryotic DNA replication initiates from multiple locations along each chromosome. These start sites, termed origins of DNA replication, are defined by primary sequence, essential initiator proteins, and local chromatin structure (Leonard and Méchali 2013). In S. cerevisiae, start sites of DNA replication are marked by the ARS consensus sequence (ACS) (Broach et al. 1983; Marahrens and Stillman 1992), a degenerate T-rich motif bound by the heterohexameric origin recognition complex (ORC) (Bell and Stillman 1992). ORC associates with replication origins throughout the cell cycle (Diffley et al. 1994). During G1 and in conjunction with Cdc6 and Cdt1, ORC recruits the Mcm2-7 helicase to initiation sites, leading to the for-
mation of the pre-replicative complex (pre-RC) (Bell and Kaguni 2013). Once the pre-RC is assembled, the origin is licensed for potential activation in S phase (Remus and Diffley 2009). Origin activation is temporally regulated, and each origin is characterized by an inherent initiation efficiency (McGuffee et al. 2013) and S-phase activation time (Raghuraman et al. 2001; Yabuki et al. 2002; Müller et al. 2014). Although the precise mechanisms regulating the selection and activation of specific origins in the genome are poorly understood, they are thought to be modulated by epigenetic features and the local chromatin environment (Rhind and Gilbert 2013).

All DNA-templated processes (e.g., transcription, replication, recombination, repair, etc.) occur in the context of the surrounding chromatin environment. Chromatin features such as nucleosome positioning can restrict access to particular DNA sequences, limiting which sites are available for protein binding (Bai and Morozov 2010). This likely contributes to the relatively small number of ~220–400 ORC-binding sites in the *S. cerevisiae* genome despite nearly 10,000 high quality ACS motif matches (Breier et al. 2004; Xu et al. 2006; Eaton et al. 2010). Classic experiments at the *ARS1* replication origin demonstrated that artificially positioning a nucleosome over the ACS impaired origin function (Simpson 1990), suggesting that nucleosomes govern ORC accessibility to the ACS. More recent high-throughput nucleosome positioning assays have established that well-positioned nucleosomes flanking the ACS are a conserved feature of replication origins throughout the genome (Eaton et al. 2010; Berbenetz et al. 2010). In addition, the flanking nucleosome positions may also play an active role in pre-RC formation, as expanding the native nucleosome-free region (NFR) at *ARS1* significantly limited pre-RC assembly (Lipford and Bell 2001). These findings demonstrated that chromatin structure significantly impacts ORC binding and pre-RC assembly.

The biochemical cascade of events leading to origin selection and ultimately activation are becoming increasingly well-understood. ORC was initially identified as a
biochemical activity that ‘footprinted’ the ACS of \textit{ARS1} by DNase I hypersensitivity (Bell and Stillman 1992). An extension of this G2 ORC footprint in G1 suggested that pre-RC assembly alters the protein-DNA occupancy at the origin (Diffley \textit{et al.} 1994; Perkins and Diffley 1998; Speck \textit{et al.} 2005). Consistent with this change in the molecular architecture of the origin, cryo-electron microscopy (cryo-EM) studies have also revealed a dramatic Cdc6-induced change in the conformation of ORC on template DNA (Sun \textit{et al.} 2012). \textit{In vitro} pre-RC assembly experiments demonstrated the capability of loading multiple Mcm2-7 double hexamers that can passively translocate along template DNA prior to activation in S phase (Bowers \textit{et al.} 2004; Evrin \textit{et al.} 2009; Remus \textit{et al.} 2009). Despite our mechanistic understanding of pre-RC assembly and initiation on template DNA \textit{in vitro} (Speck \textit{et al.} 2005; Heller \textit{et al.} 2011; Frigola \textit{et al.} 2013; Sun \textit{et al.} 2013; Fernández-Cid \textit{et al.} 2013), we have comparably little knowledge of how these events are regulated at specific origins in the chromosomal context.

Recent genome-wide datasets, including ORC and Mcm2-7 chromatin immuno-precipitation (ChIP) assays (Wyrick \textit{et al.} 2001; Xu \textit{et al.} 2006; Eaton \textit{et al.} 2010; Szilard \textit{et al.} 2010), replication timing experiments (Raghuraman \textit{et al.} 2001; Yabuki \textit{et al.} 2002; Müller \textit{et al.} 2014), and recovery of nascent DNA replication intermediates (Feng \textit{et al.} 2006; Crabbé \textit{et al.} 2010; McGuffee \textit{et al.} 2013), have provided a wealth of critical information about the precise location and efficiency of individual replication origins across the genome. However, the rules and chromatin features that govern origin selection and regulation remain elusive, including the following: (1) Where and how many Mcm2-7 complexes are loaded in relation to ORC-binding sites? (2) Do all origins share the same molecular architecture? (3) How do origin-proximal transcription factors impact origin architecture and regulation? (4) How does origin architecture change throughout the cell cycle? (5) Are specific chromatin features associated with origin efficiency and activation time? Although some
of these questions can potentially be addressed biochemically at defined origins \textit{in vitro}, there will undoubtedly be origin-specific differences that are dependent on the local chromatin environment.

To better understand how the chromatin architecture at individual replication origins impacts origin function, we used micrococcal nuclease (MNase) mapping to comprehensively “footprint” the protein-DNA occupancy across the \textit{S. cerevisiae} genome at multiple points in the cell cycle (Kent \textit{et al.} 2011; Henikoff \textit{et al.} 2011). Using this approach, we are able to identify, at nucleotide resolution, DNA fragments protected by nucleosomes, transcription factors, and key components of the pre-RC. Importantly, this technique not only provides significantly higher resolution than other approaches, like ChIP, but it is also factor-agnostic and does not rely on enrichment with specific antibodies. We identified 269 origins with an ORC-dependent chromatin footprint. These ORC-dependent footprints are specifically extended during pre-RC assembly in G1 and agreed with previous protection estimates generated from DNase I footprinting at select loci (Diffley \textit{et al.} 1994; Perkins and Diffley 1998; Speck \textit{et al.} 2005). Not all origins exhibited the same chromatin architecture; a distinct class of inefficient origins yielded a cell-cycle–dependent footprint only in G1, suggesting a transient ORC association at these origins until stabilization in G1. Chromatin remodeling of the nucleosomes flanking the ACS in G1 was a feature of efficient origins. Integrating the nucleotide-resolution “footprints” of origin architecture with ORC and Mcm2-7 ChIP-seq provided new insights into Mcm2-7 loading \textit{in vivo}. Each origin assembles an Mcm2-7 complex either upstream of or downstream from ORC, proximal to the flanking nucleosome, suggesting that, unlike \textit{in vitro}, only one Mcm2-7 double hexamer is loaded per origin. In summary, MNase mapping provides a critical independent validation of prior ChIP-seq and origin mapping experiments, and also yields new mechanistic insights into pre-RC assembly, chromatin dynamics, and origin regulation genome-wide.
3.2 Results

3.2.1 Chromatin footprinting replication origins genome-wide

To interrogate the origin chromatin structure at high resolution, we used a MNase digestion assay that reveals protein occupancy across the genome (Kent et al. 2011; Henikoff et al. 2011). Protein-bound DNA regions are protected from MNase digestion, and DNA fragments >25 base pairs (bp) can be recovered for high-throughput paired-end sequencing. We were able to identify distinct regions of DNA protection by nucleosomes (~150 bp) and smaller (25- to 120-bp) protein-DNA interactions (e.g., transcription factors and ORC). MNase mapping provides a comprehensive and unprecedented nucleotide-resolution view of chromatin structure at individual origins throughout the genome.

We first profiled the ORC-dependent chromatin architecture at DNA replication origins in G2, prior to G1 pre-RC assembly. We arrested cells at the G2/M transition by treatment with nocodazole and subjected these cells to MNase mapping. Chromatin architecture was then visualized at individual replication origins by plotting the chromosomal position of the recovered protected reads as a function of their fragment length. For example, at ARS107, which is situated in the intergenic region between PSK1 and TPD3 on chromosome I, we detected well-positioned nucleosomes, as evidenced by periodic clusters of 150- to 175-bp fragments (Fig. 3.1). In addition, we also observed two smaller protection regions (25–120 bp) that indicate specific nonnucleosomal protein-DNA interactions, mapping precisely to the annotated motifs for the ACS and the transcription factor Abf1p. Thus, we were able to detect distinct DNA-binding events in a comprehensive and factor-agnostic manner.

To confirm that the 25- to 120-bp fragments at the ACS of ARS107 originated from and were dependent on ORC occupancy, we evaluated the chromatin architecture in the temperature-sensitive orc1-161 mutant. In the orc1-161 mutant, ORC
can no longer associate with DNA at the nonpermissive temperature (37°C) (Aparicio et al. 1997). MNase mapping revealed that the chromatin protection footprint at the ACS of ARS107 was completely lost (Fig. 3.1, bottom panel); in contrast, the chromatin occupancy at the annotated Abf1-binding site was unaffected. Loss of ORC at the ACS also resulted in nucleosome encroachment over the replication
Figure 3.2: Presence of ORC-dependent footprint at replication origins genome-wide. Aggregate chromatin profiles from a previously characterized replication origin dataset (ORC-ACS) (Eaton et al. 2010), with all origins centered on and oriented by the ACS. The left and middle panels are derived from WT andorc1-161 MNase digestions, respectively. The right panel represents a merged overlay of WT (red) andorc1-161 (green).

This chromatin reorganization impacted the nucleosome positioning within thePSK1gene body but not theTPD3gene, likely due to the barrier effects ofAbf1p(Mavrich et al. 2008a; Zhang et al. 2009). Therefore, MNase mapping can capture dynamic chromatin changes that occur when protein-DNA binding is altered.

The ORC-dependent chromatin architecture we observed atARS107was representative of origins throughout the genome. We generated aggregate wild-type andorc1-161 chromatin occupancy profiles from replication origins identified in a prior ORC ChIP-seq study (Eaton et al. 2010) by aligning each origin to the 5’ position of the ACS motif (Fig. 3.2). To ensure directional consistency, we oriented the ACS at every replication origin relative to the T-rich strand. In describing chromatin features relative to the ACS, upstream and downstream correspond to the 5’ and 3’ side of the T-rich ACS strand, respectively. As previously described in wild type (Eaton et al. 2010), MNase mapping of origins revealed well-positioned nucleosomes around
Figure 3.3: *orc1-161* chromatin differs specifically at replication origins. (A) Nucleosome profiles surrounding yeast transcription start sites (TSS) (Xu et al. 2009) for WT and *orc1-161* digested chromatin. The small fragment (25–120 bp) footprint occupancy for WT and *orc1-161* at (B) Abf1-binding sites (MacIsaac et al. 2006) and (C) replication origins.

Interestingly, the bulk of the DNA protection from small fragments occurred in the middle of the NFR, suggesting that the ORC footprint extends downstream of the ACS. In contrast, DNA protection at the ACS was completely lost in the *orc1-161* mutant at the nonpermissive temperature (Fig. 3.2, middle panel). Similar to *ARS107*, loss of ORC resulted in the encroachment of nucleosomes over the ACS, highlighted by a high concentration of 150- to 175-bp reads closer to the ACS in the *orc1-161* mutant (Fig. 3.2, right panel). Importantly, we did not observe any significant differences in protection of DNA features outside of replication origins (Fig. 3.3).
Figure 3.4: ORC only associates with a subset of putative origins. Small fragments (25–120 bp) surrounding the ACS at each of the 798 putative (exhibiting ARS function on a plasmid) OriDB origins (Nieduszynski et al. 2007) were condensed into single footprint signals for (A) WT and (B) orc1-161. Each row in the heatmap represents an individual origin centered on and oriented by the ACS. (C) The orc1-161 small fragment signal was subtracted from the WT small fragment signal to yield a difference heatmap. Higher signal in WT or orc1-161 is represented by red or green intensity, respectively. 269 origins (top subset) contain an ORC-dependent footprint. (D) ORC ChIP-seq data from asynchronous cells largely overlapped with replication origins containing an ORC-dependent footprint (top subset).
Figure 3.5: Transient ORC association at a subset of replication origins. Heatmap of small fragment (25–120 bp) footprint occupancy surrounding the ACS at each putative OriDB origin (Nieduszynski et al. 2007) for G1-arrested (left panel) and G2-arrested (right panel) cells. ORC-dependent footprint origins displayed an increased occupancy and wider protection region in G1 (G1 & G2 Footprint). A separate origin class exhibited protection of the ACS region specifically in G1 (G1-Only Footprint).

The MNase assay provides an independent and high-resolution method to assess the chromatin occupancy at all potential replication origins genome-wide. We compared the G2 protection of the ACS in wild type (Fig. 3.4A) and orc1-161 (Fig. 3.4B) at nearly 800 putative origins that exhibited ARS function in a plasmid-based as-
Figure 3.6: Replication origins increase protein occupancy in G1. The G1 small fragment (25–120 bp) occupancy for G1 WT (red), G1 cdc6-1 (purple), and G2 WT (green) for (A) replication origins and (B) Abf1-binding sites (MacIsaac et al. 2006).

3.2.2 ORC only associates stably with a subset of functional origins in G2

As the cell cycle progresses through G1, ORC, together with Cdc6 and Cdt1, directs the loading of the Mcm2-7 helicase at replication sites to form the pre-RC (Bell and Kaguni 2013). Although pre-RC formation has been extensively studied in vitro (Speck et al. 2005; Frigola et al. 2013; Sun et al. 2013; Fernández-Cid et al. 2013;
Figure 3.7: *cdc6-1* is defective in Mcm2-7 loading. ORC (top) and Mcm2-7 (bottom) occupancy from WT and *cdc6-1* strains were determined by ChIP in G1 and evaluated by quantitative-PCR. Fold ChIP enrichment was normalized to *ARS1* WT.

Heller et al. (2011), footprinting studies have been limited to only a few origins (Diffley et al. 1994; Perkins and Diffley 1998; Speck et al. 2005). It remains unclear how the local chromatin architecture and origin-specific protein-DNA interactions impact pre-RC assembly and subsequent DNA replication initiation.

To address these questions, we evaluated the origin-specific chromatin architecture during ORC binding in G2/M (nocodazole arrest) and following pre-RC assembly in G1 (α-factor arrest). We observed a small but significant expansion in the ORC-dependent footprint from G2 to G1 (Figs. 3.5 and 3.6A). Importantly, this footprint expansion was specific to replication origins, as other transcription-factor–binding sites did not yield any protection differences between G2 and G1 (Fig. 3.6B). To test if this expansion was dependent on pre-RC assembly, we used a temperature-sensitive allele of *CDC6* (*cdc6-1*) (Hartwell et al. 1973) to abrogate pre-RC assem-
Figure 3.8: Transient ORC association does not impair Mcm2-7 loading. (A) Distribution of asynchronous ORC ChIP-seq enrichment for each putative origin. ORC ChIP-seq signal was significantly greater in the G1 & G2 Footprint class compared to either the G1-Only Footprint or No Footprint classes (Wilcoxon test, $P < 6.80 \times 10^{-13}$ and $P < 4.81 \times 10^{-43}$, respectively). A smaller yet significant difference was also detected in the G1-Only Footprint class relative to the No Footprint class ($P < 1.05 \times 10^{-7}$). (B) Distribution of G1 Mcm2-7 ChIP-seq enrichment for each putative origin. G1 & G2 Footprint and G1-Only Footprint classes showed no significant difference in Mcm2-7 signal (Wilcoxon test, $P < 0.035$), but each contained greater Mcm2-7 signal than the No Footprint class ($P < 3.48 \times 10^{-43}$ and $P < 3.93 \times 10^{-28}$, respectively).

In the absence of Cdc6 activity, we found that the expansion of the G1 footprint at origins was lost, and the chromatin more closely resembled the G2 state (Fig. 3.6A). The expanded G1 footprint may reflect increased protection conferred by the ORC-Cdc6 complex or the Mcm2-7 double hexamer.

A subset of origins exhibited a G1 chromatin occupancy at the ACS yet lacked a G2 ORC-dependent footprint (Fig. 3.5). The absence of a G2 ORC-dependent footprint suggested that ORC either binds transiently in G2 or does not associate with these origins until G1 and pre-RC assembly. Not surprisingly, ORC ChIP-seq signal in asynchronous cells was substantially less at these origins relative to the sites
Figure 3.9: ACS sequence is not predictive of occupancy differences between the G1 & G2 Footprint and G1-Only Footprint classes. Position-weight matrices were derived from the ACS sequences at (A) G1 & G2 Footprint, (B) G1-Only Footprint, or (C) No Footprint replication origins.

exhibiting a G2 ORC-dependent footprint (Fig. 3.8A). Impaired ORC binding is not likely caused by a weak ACS motif, as both origin classes yielded nearly identical sequence motifs (Fig. 3.9).

We next asked whether the lack of an ORC-dependent footprint in G2 had any functional consequences. To assess origin function, we used the distribution of Okazaki fragments (McGuffee et al. 2013) and BrdU incorporation during a hydroxyurea arrest as proxies for origin efficiency and origin activation time, respectively. We found that origins not harboring a G2 ORC-dependent footprint are much more likely to be dormant, inefficient, and late-firing (Fig. 3.10). However, this defect
in origin activation does not appear to be due to diminished pre-RC assembly, as there is no significant reduction in Mcm2-7 loading as determined by ChIP-seq enrichment for this origin class (Fig. 3.8B). Together, these data suggest that stable ORC association in G2 is a determinant for efficient and early-activating replication origins.

3.2.3 Nucleosome dynamics in G1 promote origin activation

A defining feature of replication origins is a well-established NFR flanked by nucleosomes asymmetrically-positioned around the ACS (Eaton et al. 2010; Berbenetz et al.)
Disruption of this optimal nucleosome positioning, such as artificial movement of the downstream nucleosome at ARS1, adversely affects replication function (Lipford and Bell 2001). Although the sequences at the replication origin are inherently nucleosome-disfavoring, ORC binding at the ACS and an ATP-dependent chromatin remodeling activity are necessary and sufficient to establish this nucleosome organization in vitro (Eaton et al. 2010). However, it is currently unknown whether pre-RC assembly or replication initiation requires additional chromatin remodeling.

To address this question, we compared the G1 and G2 flanking nucleosome positions from all origins harboring a chromatin footprint in G1. We constructed G1 and G2 nucleosome profiles from the midpoint positions of all 150- to 175-bp DNA fragments surrounding replication origins. To focus on cell-cycle—dependent differences in nucleosome positioning, we plotted the log ratio of G1 to G2 nucleosome
Figure 3.12: Nucleosome movement is dependent on pre-RC formation. (A) Heatmap representing differences between nucleosome density in G1 cdc6-1 (red) and G2 WT (green), ordered as in Figure 3.11A. (B) Average G1 cdc6-1 and G2 WT nucleosome densities were generated for each nucleosome class.

density as a heatmap (Fig. 3.11A). We observed three different cell-cycle–dependent nucleosome configurations. We found that 41% of the origins maintained static nucleosome positioning between G1 and G2; in contrast, the other two origin configurations exhibited expanded NFRs in G1 by shifting either the upstream (26%) or downstream (33%) nucleosome but not both. Average G1 and G2 nucleosome density plots exhibited well-positioned nucleosomes within each class (Fig. 3.11B). We conclude that a subset of origin-flanking nucleosomes undergo cell-cycle–dependent chromatin remodeling immediately prior to or during pre-RC assembly.

We hypothesized that the chromatin remodeling events observed at a subset of origins may be dependent on pre-RC assembly. To test this hypothesis, we blocked pre-RC formation by utilizing a temperature-sensitive cdc6-1 mutant and assayed nucleosome positioning in α-factor G1-arrested cells at the restrictive temperature. In the absence of pre-RC formation, all origins now maintained their consensus G2
Figure 3.13: Dynamic nucleosome position predicts ORC-dependent footprint expansion direction. Average small fragment (25–120 bp) footprint occupancy surrounding the ACS for each class of nucleosome positioning in G1 WT (red), G1 cdc6-1 (purple), and G2 WT (green).

nucleosome positions (Fig. 3.12), thus establishing a link between pre-RC assembly and nucleosome remodeling at a subset of replication origins.

Presumably, the chromatin remodeling and subsequent enlargement of the NFR directs pre-RC assembly and Mcm2-7 helicase loading. We re-examined the expansion of the ORC-dependent footprint in G1 for each class of observed nucleosome movements: upstream, downstream, and static. We found that the footprint expansion at individual origins was predictive for each class of nucleosome movements (Fig. 3.13). Specifically, an upstream expansion of the origin-specific footprint was associated with the movement of the upstream nucleosome, and vice versa for the downstream expansion. Not surprisingly, limited expansion of the footprint was observed for the static class of nucleosome movements. Finally, as with the nucleosome remodeling, the footprint expansion in G1 was dependent on Cdc6 activity.
Figure 3.14: Cell-cycle–dependent nucleosome dynamics promote origin efficiency. (A) Origin activation and efficiency for each class of nucleosome positioning. Both the upstream and downstream shifting nucleosome classes were more efficient than the static nucleosome class (Wilcoxon test, upstream vs. static, $P < 1.42 \times 10^{-4}$; downstream vs. static, $P < 2.03 \times 10^{-7}$; upstream vs. downstream, $P < 0.134$) and were earlier activating ($\chi^2$ test, upstream vs. static, $P < 0.0206$; downstream vs. static, $P < 1.04 \times 10^{-4}$; upstream vs. downstream, $P < 0.474$). (B) Distribution of G1 Mcm2-7 ChIP-seq enrichment across each nucleosome class. Mcm2-7 ChIP-seq signal was significantly less in the static class relative to the dynamic nucleosome classes (Wilcoxon test, upstream vs. static, $P < 5.99 \times 10^{-4}$; downstream vs. static, $P < 7.95 \times 10^{-7}$). The upstream and downstream nucleosome classes did not significantly differ in Mcm2-7 occupancy ($P < 0.136$).

The expanded G1 footprint may reflect increased protection conferred by the ORC-Cdc6 complex, which in vitro extends the ORC footprint by $\sim 30$ bp (Speck et al. 2005), or the Mcm2-7 double hexamer, whose 230 Å central channel protects $\sim 68$ bp of DNA (Remus et al. 2009). As we observed a $\sim 25$ bp average upstream or downstream expansion (Fig. 3.13), we have most likely recovered the ORC-Cdc6 complex. We speculate that the Mcm2-7 complex, which can freely slide on DNA templates (Remus et al. 2009; Evrin et al. 2009), translocates off the MNase-digested DNA due to the limited efficiency of formaldehyde crosslinking (Schmiedeberg et al. 81).
Is nucleosome remodeling important for origin function? We examined origin efficiency and activation time for the three classes of nucleosome movement (upstream, downstream, and static) (Fig. 3.14A). We found that origins that underwent nucleosome remodeling in G1 were significantly more efficient. This increased origin activation may be explained by additional pre-RC assembly in the presence of dynamic nucleosomes. We examined Mcm2-7 ChIP-seq enrichment within each replication origin class (Fig. 3.14B). We found that Mcm2-7 loading was significantly impaired in the static class, although clearly not reduced to background levels. These results suggest that G1 nucleosome movement promotes origin activation at the pre-RC assembly step; however, given that Mcm2-7 levels are not reduced to background levels, we cannot rule out the possibility that the nucleosome remodeling facilitates steps subsequent to pre-RC assembly, including origin unwinding, helicase activation, or DNA polymerase recruitment.

3.2.4 Origin-flanking nucleosomes restrict Mcm2-7 loading

Our chromatin structure analysis using MNase digestion revealed cell-cycle–dependent changes in origin architecture, but did not provide insight into where the Mcm2-7 complex was loaded relative to ORC. Given the large NFR downstream of the ACS, we hypothesized that Mcm2-7 double-hexamer loading is confined to the unoccupied gap between ORC and the downstream nucleosome. To test this hypothesis, we used ChIP-seq to precisely localize ORC and Mcm2-7 relative to the chromatin features we characterized using MNase mapping (Fig. 3.15). We first identified the peaks from ORC (Fig. 3.15, green) or Mcm2-7 (Fig. 3.15, purple) ChIP-seq experiments. Relative to the chromatin environment revealed by MNase mapping (Fig. 3.15, gray), we found that the ORC ChIP-seq peaks were concentrated in the NFR and coincided with the ACS. In contrast, we found that the Mcm2-7 ChIP-seq peaks were located
Figure 3.15: Mcm2-7 loading coincides with the flanking nucleosomes. Locations of ORC (green) and Mcm2-7 (purple) ChIP-seq peaks relative to the G1 MNase chromatin data (gray) surrounding origins. For each of the 398 origins with a G1 footprint, 257 asynchronous ORC and 332 G1 Mcm2-7 peaks were identified. The ChIP-seq peak position relative to the ACS was plotted on the x-axis, and random noise was applied along the y-axis to spread the data points.

either upstream of or downstream from the ORC footprint and appeared to overlap with the flanking nucleosome positions (only one peak was associated with each origin). The apparent occupancy of the same DNA sequence by both the histone octamer and Mcm2-7 suggested the following models: (1) a dynamic competition between Mcm2-7 and nucleosomes resulted in stochastic occupancy of the same position. (2) The Mcm2-7 complex protected an ~150 bp region of DNA mimicking a histone octamer. (3) The Mcm2-7 complex was immediately adjacent to or in close contact with the nucleosome, and they co-immunoprecipitated as a complex.

We first considered stochastic model 1, in which the origin-flanking DNA is oc-
Figure 3.16: No evidence of nucleosome eviction at the Mcm2-7 loading position. Average nucleosome densities for the (A) upstream and (B) downstream Mcm2-7 loading classes show no substantial cell-cycle-dependent nucleosome occupancy differences. Nucleosome density is not greatly lost between G1 and G2 at the Mcm2-7 loading position, nor is nucleosome occupancy significantly different between either flanking nucleosome. Nucleosome, red ellipse; Mcm2-7, purple oval.

occupied by either a histone octamer or the Mcm2-7 complex. If Mcm2-7 loading resulted in nucleosome eviction, then replacement of a nucleosome by a Mcm2-7 double hexamer should substantially reduce the G1 nucleosome occupancy. We classified origins into two groups, upstream and downstream, based on the Mcm2-7 ChIP-seq enrichment at each flanking nucleosome and compared the G1 and G2 nucleosome occupancy. However, we detected no cell-cycle-dependent nucleosome density differences at the Mcm2-7 enrichment location (Fig. 3.16). Further, both origin-flanking nucleosomes exhibited similar occupancy levels within each group. Combined, these results fail to support a model of stochastic nucleosome occupancy at Mcm2-7 binding locations.

We next considered model 2, in which the Mcm2-7 complex was protecting an ~150 bp fragment of DNA analogous to the protection conferred by a histone oc-
Figure 3.17: Nucleosome occupancy explains protected fragments at Mcm2-7 loading position. Fragment length distribution of origin-flanking regions for upstream (left) and downstream (right) Mcm2-7 occupancy classes. Mcm2-7, purple oval.

tamer. However, it seemed unlikely that a double hexamer of Mcm2-7 whose 230 Å central channel is predicted to only protect ~68 bp of DNA (Remus et al. 2009) would be sufficient to generate an ~150 bp occupancy footprint. Furthermore, both upstream and downstream Mcm2-7 ChIP enrichment groups revealed no differences in the length distributions of MNase-protected fragments (Fig. 3.17), strongly suggesting that the MNase-protected occupancy is due to a histone octamer.

Finally, we considered model 3, in which the Mcm2-7 double hexamer may be immediately adjacent to or in complex with the flanking nucleosomes. We analyzed the distribution of the forward (upstream) and reverse (downstream) reads that are generated from every sequenced immunoprecipitated fragment within each Mcm2-7
ChIP-seq enrichment group (Fig. 3.18A). The forward (Fig. 3.18A, red) and reverse (Fig. 3.18A, green) read positions demarcate the ChIP enrichment boundaries and precisely identify the protein-binding region. When we plotted the forward and reverse reads for the ORC-associated fragments (Fig. 3.18B, middle left panel), regardless of whether the origin loaded Mcm2-7 upstream or downstream, we found that the peaks of reverse (Fig. 3.18B, green) reads were located in the NFR downstream from the ACS (Fig. 3.18B, top left and bottom left panels). The forward ORC reads (Fig. 3.18B, red) exhibited a bimodal peak indicating that the fragmentation upstream of ORC occurred either between ORC and the adjacent nucleosome or upstream of the nucleosome. As this pattern was only detected at the upstream nucleosome, it may indicate an interaction between ORC and the adjacent nucleosome, potentially mediated by the BAH-domain of Orc1p (Onishi et al. 2007; Müller et al. 2010). Unlike with ORC, no putative nucleosome interactions were observed for the transcription factor Abf1p by ChIP-seq (Fig. 3.19).

In contrast, the forward and reverse peak positions of the Mcm2-7 ChIP-seq reads encompassed the flanking nucleosomes (Fig. 3.18B, middle right panel). In the cases where Mcm2-7 was loaded upstream of the origin, we found that the forward reads (Fig. 3.18B, red) coincided with the left hand edge of the nucleosome and that the reverse reads (Fig. 3.18B, green) were within the NFR (Fig. 3.18B, top right panel). Similarly, origins loading Mcm2-7 downstream yielded a complementary pattern in the forward and reverse read distributions flanking the downstream nucleosome (Fig. 3.18B, bottom right panel). Importantly, the total distance between the peaks of forward and reverse reads was ~210 bp, almost exactly the expected width of one nucleosome (~150 bp) and Mcm2-7 double hexamer (~68 bp). Finally, we found no differences in origin activation or efficiency that could be attributed to Mcm2-7 loading upstream of or downstream from the ACS (Fig. 3.20). Similarly, we were unable to link the Mcm2-7 loading side with the directional Cdc6-dependent
Figure 3.18: Mcm2-7 loads either upstream or downstream of ORC. (A) Hypothetical distribution of ChIP-seq forward (red) and reverse (green) reads representing chromatin fragmentation patterns around a specific DNA-binding factor. The most likely protein-binding region can be inferred as the area between the two read distributions. (B) Analysis of chromatin fragmentation patterns resulting from ORC (left) and Mcm2-7 (right) ChIP-seq. 389 origins with sufficient Mcm2-7 ChIP-seq signal were grouped into two classes (upstream and downstream) based on the enrichment signal relative to the origin-flanking nucleosomes. Heatmaps (separated by class) represent the distributions of forward (red) and reverse (green) reads at each origin relative to the ACS for ORC (left) and Mcm2-7 (right). Origins have the same ordering in both heatmaps. Top and bottom panels represent the average forward and reverse ORC and Mcm2-7 ChIP-seq densities within each class. Consensus fragmentation boundaries are designated by red and green dotted lines for forward and reverse reads, respectively. The average nucleosome density (blue) for each class is superimposed on each plot, and inferred nucleosome positions are depicted as red shaded ovals. ORC, green rectangle; Mcm2-7, purple oval.
Abf1p does not physically interact with adjacent nucleosomes. Analysis of chromatin fragmentation patterns resulting from Abf1p ChIP-seq. A heatmap representation of forward (red) and reverse (green) read distributions at each Abf1-binding site (MacIsaac et al. 2006). Top panel represents the average forward and reverse ChIP-seq densities, and consensus fragmentation boundaries are designated by red and green dotted lines, respectively. The average nucleosome density (blue) is superimposed, along with inferred nucleosome positions (red shaded ovals) and Abf1 (green rectangle).

ORC footprint expansion and nucleosome shift. We conclude that origin licensing can yield at least two pre-RC configurations in which Mcm2-7 has been preferentially loaded adjacent to, and likely in complex with, either the upstream or downstream nucleosome.
Figure 3.20: Mcm2-7 loading at either flanking nucleosome can catalyze replication initiation. Origin activation and efficiency for each Mcm2-7 occupancy class. No origin efficiency differences (Wilcoxon test, $P < 0.4664$) or activation time differences ($\chi^2$ test, $P < 0.055$) were detected.

### 3.3 Discussion

Utilizing MNase mapping, we comprehensively profiled the cell-cycle–dependent local chromatin environment surrounding DNA replication origins. This assay maps DNA-binding protein occupancy across the genome in a cost-effective and factor-agnostic manner. We precisely identified 269 ORC-dependent footprints in G2 and found that only a subset of all putative origins (identified in a plasmid-based assay) harbored an ORC-dependent footprint *in vivo*. We also found that differences in origin efficiency were associated with cell-cycle–dependent changes in both origin chromatin architecture and ORC association. Chromatin remodeling of the flanking
nucleosomes between G2 and G1 was linked to active and efficient replication origins and was dependent on Cdc6. Finally, integration of the origin chromatin occupancy profiles with ORC and Mcm2-7 ChIP-seq revealed that the Mcm2-7 double hexamer preferentially loads either upstream of or downstream from ORC, adjacent to the flanking nucleosome. These data provide new insights into how local chromatin architecture modulates origin selection and function.

Genome-wide approaches based on ChIP-chip and ChIP-seq targeted against pre-RC components have been used by multiple groups to pinpoint origins in their chromosomal context (Wyrick et al. 2001; Xu et al. 2006; Eaton et al. 2010; Szilard et al. 2010). While ChIP has the potential for increased sensitivity (due to the biochemical enrichment by immunoprecipitation), it is also susceptible to nonspecific interactions (Teytelman et al. 2013) and epitope masking (Aparicio et al. 1997), which may result in the detection of false positives and negatives, respectively. Indeed, the number of ORC binding sites in yeast varies from conservative estimates of 220 to potentially >400, and a number of putative origins contain Mcm2-7 ChIP enrichment without any ORC ChIP signal (Wyrick et al. 2001; Xu et al. 2006; Eaton et al. 2010). In comparison with our previously identified 253 conservative ORC ChIP-seq peaks (Eaton et al. 2010), 179 are present in our G1 & G2 Footprint dataset (70.8% overlap). The 74 ORC sites identified by the ChIP assay and not MNase footprinting may represent weak and/or transient binding interactions, or may represent false positives from nonspecific interactions (Teytelman et al. 2013). Conversely, the MNase assay detected 90 ORC-dependent footprints in putative origins that were missed by the ChIP-seq assay. As these origins all harbored an ORC-dependent footprint, it is consistent with prior reports indicating that a significant fraction of ORC-binding sites may be missed by ChIP-seq due to epitope masking (Wyrick et al. 2001; Xu et al. 2006). In summary, although the MNase-based assay may miss weak or transient interactions, it provides increased resolution compared to ChIP approaches, is
not sensitive to epitope masking, and serves as an important independent validation of origin location throughout the genome.

A subset of origins had no discernable ORC-dependent occupancy in G2, yet exhibited a protected footprint in G1 (Fig. 3.5). Since the MNase chromatin assay may not detect transient interactions, we propose that ORC may only weakly or dynamically associate with these origins until pre-RC assembly in G1 stabilizes ORC binding. These origins have limited G2 ORC residence, supported by a significantly reduced ORC ChIP signal from an asynchronous cell population (Fig. 3.8A). Origins that failed to associate with ORC in G2 had a dramatically reduced origin efficiency (Fig. 3.10), reminiscent of prior observations in *Schizosaccharomyces pombe* (Wu and Nurse 2009), which linked time of ORC association in mitosis with origin function. We propose that constitutive ORC association with DNA is one determinant of origin efficiency, and as cells enter G1, ORC-associated origins will immediately commence with pre-RC assembly; however, pre-RC formation will likely be delayed for those origins lacking a stable ORC-DNA interaction prior to mitosis. Even though the final levels of Mcm2-7 loading appear unaffected (Fig. 3.8B), this pre-RC assembly delay possibly handicaps recruitment of Cdc45, Dbf4, and other limiting replication initiation factors (Aparicio *et al.* 1997; Pasero *et al.* 1999; Mantiero *et al.* 2011; Tanaka *et al.* 2011).

Pre-RC formation in G1 leads to Cdc6-dependent chromatin remodeling events that subsequently impact origin efficiency. We identified three cell-cycle-dependent nucleosome configurations at replication origins in G1 (Fig. 3.11). The first two classes exhibited either upstream or downstream (but not both) nucleosome movement away from the ACS-oriented origin, while a third comparably inefficient static class did not reposition either flanking nucleosome. We considered the straightforward hypothesis that G1 nucleosome movement increases the origin NFR and thus provides more space for Mcm2-7 loading. However, even though the static class
had significantly reduced Mcm2-7 levels, they were still above background levels of Mcm2-7 loading (Fig. 3.14B). We propose that in addition to promoting Mcm2-7 loading, origin nucleosome remodeling may also establish a more favorable chromatin environment for subsequent replication initiation events, including Cdc45 and GINS recruitment (Ilves et al. 2010), DNA unwinding, or DNA polymerase loading. Since precise nucleosome positioning is important for Mcm2-7 assembly (Lipford and Bell 2001), nucleosome displacement is possibly regulated by specific chromatin remodelers like Ino80 or Isw2. These remodelers have already been associated with replication forks (Shimada et al. 2008; Vincent et al. 2008), but our results suggest a potentially new role for remodelers at replication origins.

Many in vitro pre-RC assembly studies have focused on the molecular mechanisms regulating Mcm2-7 helicase loading. Since the ACS is asymmetrically located closer to the upstream nucleosome, several models posit that ORC facilitates Mcm2-7 loading in the downstream NFR (Eaton et al. 2010). We examined the structural chromatin changes in origin architecture that occur during pre-RC assembly. First, we identified a ~25 bp Cdc6-dependent upstream or downstream extension of the origin-specific footprint in G1, accompanied by a directional expansion of the NFR (Fig. 3.13). Previous in vitro DNase I hypersensitivity mapping at individual origins estimated a similar ORC footprint expansion induced by Cdc6 binding (Speck et al. 2005). A recent ORC-Cdc6 cryo-EM study suggested that ORC adopts a crescent structure along the DNA, with Cdc6 bridging the gap between the ends of the crescent where DNA both enters and exits the ORC-Cdc6 structure (Sun et al. 2012). We propose that the ORC-Cdc6 footprint extension direction is driven by chromatin or sequence-specific interactions that influence MNase accessibility to the DNA at the interface of the ORC-Cdc6 structure.

By integrating ORC and Mcm2-7 ChIP-seq data with our high-resolution chromatin profiles, we interrogated the exact location of pre-RC assembly at replication
origins across the genome (Figs. 3.15 and 3.18). ORC, as expected, mapped to the ACS within the NFR. In contrast, the Mcm2-7 complex localized either upstream of or downstream from ORC (oriented relative to the ACS). Given the upstream or downstream expansion of the ORC-dependent footprint in G1, we expected to find that the direction of the Cdc6-mediated expansion would be predictive of the Mcm2-7 double-hexamer loading side; however, the ORC footprint expansion side was not correlated with the Mcm2-7 loading position. We next considered whether the Mcm2-7 loading mechanism could plausibly support two different pre-RC assembly orientations. Recent cryo-EM structures indicate that ORC-Cdc6 undergoes a substantial conformational change from a flat crescent structure (Sun et al. 2012) to a right-handed spiral when the Cdt1–Mcm2-7 complex initially loads (Sun et al. 2013). The DNA also undergoes a dramatic transition from wrapping inside ORC-Cdc6 to threading vertically through the ORC–Cdc6–Cdt1–Mcm2-7 (OCCM) intermediate structure. However, the mechanism reorienting the DNA remains unclear, and we speculate that the planar DNA within the ORC–Cdc6 complex can be rotated either clockwise or counterclockwise to occupy the OCCM inner channel. Depending on the rotation direction, the Mcm2-7 complex would be loaded either upstream of or downstream from the ACS. Primary sequence or topological constraint likely dictates the preferred DNA orientation within the OCCM structure. Alternatively, the preferential Mcm2-7 steady-state position might be chosen independently from the loading side; however, this would require substantial ORC turnover to permit Mcm2-7 sliding past ORC.

Prior work from the Bell laboratory demonstrated the importance of precisely positioned flanking nucleosomes facilitating Mcm2-7 loading (Lipford and Bell 2001). Interestingly, we find that Mcm2-7 appears to occupy the same region of DNA as the flanking nucleosome (Fig. 3.15). We considered several possibilities to explain the apparent simultaneous DNA occupancy of Mcm2-7 and the flanking nucleosome
(detailed in Results), but our data were most consistent with a nucleosome being pulled down together with the immunoprecipitated Mcm2-7 complex. Prior experiments have reported that certain Mcm2-7 subunits directly interact with histone H3 (Ishimi et al. 1996), and the Mcm2-7 complex mediates histone contacts with FACT and Asf1 at replication forks (Groth et al. 2007; Foltman et al. 2013). We propose that the origin-flanking nucleosomes both positionally confine and engage with Mcm2-7 in vivo to facilitate subsequent unwinding of the origin DNA in S phase.

How many Mcm2-7 double hexamers are loaded at individual origins? In theory, only one double hexamer is required at each origin, resulting in one Mcm2-7 hexamer for each replication fork. However, it is clear from in vitro experiments that it is possible to load more than one Mcm2-7 complex on template DNA (Remus et al. 2009). Similarly, early work in S. cerevisiae identified a vast excess (~20-fold) of Mcm2-7 molecules relative to ORC in whole-cell extracts (Donovan et al. 1997). Together, these data support a model in which multiple Mcm2-7 double hexamers have the potential to load at chromosomal origins in vivo. However, our data suggest an alternative model in which only one Mcm2-7 double hexamer is loaded per origin. Specifically, we observed only one Mcm2-7 ChIP-seq peak per origin, and the Mcm2-7 complex appears to be in a tight association with the origin-flanking nucleosome. Analysis of the chromatin fragmentation distributions from the Mcm2-7 ChIP-seq suggest an immunoprecipitated DNA fragment of ~210 bp, almost exactly the predicted coverage of a histone octamer (~150 bp) and one Mcm2-7 double hexamer (~68 bp). Although the most direct interpretation of our data supports only one Mcm2-7 double hexamer per origin, we cannot rule out more complex models of Mcm2-7 loading (including transient and weak interactions) that may result in multiple complexes per origin.

Elegant in vitro experiments reconstituting pre-RC assembly (Speck et al. 2005; Fernández-Cid et al. 2013; Frigola et al. 2013; Sun et al. 2013) and replication initi...
tion (Remus et al. 2009; Heller et al. 2011) have provided tremendous mechanistic insights, but assays performed on template DNA clearly lack the regulatory complexity observed at each origin in its chromosomal context. For example, unlike the situation in vivo, pre-RC assembly and initiation on template DNA are not strictly dependent on origin DNA sequences (Remus et al. 2009; Gros et al. 2014). In higher eukaryotes, there is considerable plasticity in the DNA replication programs with the selection and activation of tissue- and development-specific replication origins (Rhind and Gilbert 2013). Since the genome sequence is static, regulation of these origins is very likely driven by local changes in the chromatin environment. Consistent with this hypothesis, defining features of ORC localization in Drosophila are dynamic nucleosome turnover (Deal et al. 2010) and ATP-dependent chromatin remodelers (Eaton et al. 2011). Just as the expression of specific genes is fine-tuned in myriad ways, we expect that diverse chromatin-mediated mechanisms will govern the selection and activation of individual eukaryotic DNA replication origins.

### 3.4 Methods

#### 3.4.1 Yeast strains

All strains are in the W303 background and have the basic genotype ade2-1, ura3-1, his3-11,15, trp1-1, leu2-3,112, can1-100.

- **DMMy70**
  - MATa, ade2-1, his3-11,15, leu2-3,112, can1-100,
  - bar1Δ::TRP, UR3::BrdU-Inc

- **DMMy31**
  - MATa, ade2-1, his3-11,15, leu2-3,112, can1-100, orc1-161

- **DMMy113**
  - MATa, ade2-1, ura3-1, his3-11,15, leu2-3,112, can1-100,
  - bar1Δ::hisG, cdc6-1 (G260D)
3.4.2 Cell synchronization

Wild-type and mutant (*orc1-161, cdc6-1*) yeast strains were grown in rich medium for all experiments. Cells were grown at room temperature to OD

600 0.6. To arrest cells at G2/M, nocodazole (Sigma) was added to a final concentration of 15 µg/mL and cells were allowed to grow for 1 h at 24°C. Cultures were then shifted to the nonpermissive temperature, 37°C, for an additional hour. Cells were arrested in G1 by treatment with α-factor (GenWay) at a concentration of 50 ng/mL.

3.4.3 Chromatin preparation

Cells were crosslinked by adding formaldehyde to a final concentration of 1% at room temperature, with stirring, for 30 min. Formaldehyde was quenched by adding glycine to a final concentration of 125 mM at room temperature, with stirring, for 5 min. Cells were centrifuged at 2,000 rpm for 5 min, washed with sterile water, and resuspended with 20 mL Buffer Z (0.56 M sorbitol, 50 mM Tris pH 7.4, autoclaved). 14 µL β-ME (United States Biological) and 0.5 mL of a 10 mg/mL solution of zymolyase prepared in Buffer Z were added. Samples were incubated at 24°C for 30 min with shaking. Cells were centrifuged at 1,500 rpm for 6 min at 4°C, then resuspended in 2.5 mL NP buffer (1 M sorbitol, 50 mM NaCl, 10 mM Tris pH 7.4, 5 mM MgCl₂, 1 mM CaCl₂, autoclaved) supplemented with 0.5 mM spermidine, 0.007% β-ME, and 0.075% NP-40. Aliquots of 15 U/µL MNase (Worthington) were prepared to determine the best digestion conditions: 1.5 mL tubes contained 4 µL, 2 µL, 1 µL, 0.5 µL, 0.25 µL, 0 µL MNase, to which 400 µL of the resuspended cells were added. Samples were inverted to mix and digested on the benchtop for 20 min. 100 µL stop buffer (5% SDS, 50 mM EDTA) and proteinase K to a 0.2 mg/mL final concentration were added, the samples were inverted, and then incubated at 65°C overnight. DNA was recovered by phenol extraction and isopropanol precipitation.
3.4.4 Chromatin immunoprecipitation

50 mL cultures of wild-type cells were grown at 30°C to $O_{D_{600}}$ 0.6. For the G1-arrested samples, $\alpha$-factor (GenWay) was added to a final concentration of 50 ng/mL; for the G2-arrested samples, nocodazole was added to a final concentration of 15 µg/mL. Cells were incubated for 2 h at 30°C. The chromatin extract was prepared and immunoprecipitated as previously described (Strahl-Bolsinger et al. 1997) with the following adaptations. Sonication was performed on an Ultrasonic 150T dismembrator (Fisher) at power level 030 6x for 15 s with incubation on ice for 1 min between each sonication. Samples were then centrifuged at 13,000 rpm at 4°C for 10 min to remove cell debris. The supernatant was immunoprecipitated with 1 µL anti-ORC antibody (gift from A. Schwacha, University of Pittsburgh) for asynchronous cells or 25 µL AS1.1 antibody (anti-Mcm2-7) for the G1-arrested cells overnight, with rocking, at 4°C. 50 µL of Gamma G beads (Amersham) were added and incubated with rocking at 4°C for 6 h. DNA was recovered as previously described (MacAlpine et al. 2010).

3.4.5 Early origin activation

25 mL wild-type cells were grown at 30°C to $O_{D_{600}}$ 0.6. $\alpha$-factor was added to a final concentration of 50 ng/mL, and cells were incubated for 2 h. Cells were centrifuged, washed with sterile water twice, and resuspended in 25 mL medium containing 0.2 M hydroxyurea and 400 µg/mL 5-bromodeoxyuridine. Cells were incubated for 1.5 h and washed twice with sterile water. The cell pellet was quick frozen and stored at $-80°C$. The genomic DNA was extracted as above, and 25 µg of DNA was used in the BrdU immunoprecipitation as previously described (Ryba et al. 2011). The entire BrdU immunoprecipitated pellet was used in the Illumina library preparation.
3.4.6 Sequencing library preparation

Illumina sequencing libraries of MNase-treated DNA were prepared using 500 ng of DNA as previously described (Henikoff et al. 2011). Illumina sequencing libraries of ChIP and BrdU DNA were prepared following Illumina TruSeq protocols.

3.4.7 Sequencing read alignment to the genome

All reads were aligned to the sacCer2/R61 version of the \textit{S. cerevisiae} genome using Bowtie 0.12.7 (Langmead et al. 2009). The recovered sequences from all paired-end MNase reads were truncated to 20 bp and aligned in paired-end mode using the following Bowtie parameters: `-n 2 -l 20 -m 1 -k 1 -X 1000`. Single-end ORC, Mcm2-7, and BrdU ChIP-seq reads were aligned with the following Bowtie parameters: `-n 2 -l 30 -m 1 -k 1`.

3.4.8 Data availability

All data has been deposited in the NCBI SRA with SRA Study Accession Number SRP041314.

3.4.9 Construction of individual data signal tracks

To facilitate more consistent analysis between replicates and across conditions, aligned MNase reads were subdivided into two classes: nucleosome reads (fragment width of 150–175 bp) and small fragment reads (fragment width of 25–120 bp). Within each class, the midpoint position of every sequencing read was extracted, and a 20-bp bandwidth Gaussian kernel was constructed around each position to smooth the signal tracks. Then, individual density kernels were summed together to create an aggregate signal track across each chromosome. Next, the signal tracks for each experiment were quantile-normalized to the matched small fragment or nucleosome signal from the G2 wild-type (WT) replicate 1 experiment. Finally, an
overall density signal for each experimental condition was obtained by averaging the quantile-normalized signal from the two replicates. In total, eight separate signal tracks were created (three 150–175 bp nucleosome tracks and three 25–120 bp sub-nucleosome tracks), consisting of averaged signals across the two replicates of G1 WT, G1 cdc6-1, G2 WT, or G2 orc1-161.

A similar method was employed for the ORC and Mcm2-7 single-end ChIP-seq datasets, except that the read position was first shifted 75 bp to infer the expected midpoint location. 30 bp bandwidth Gaussian kernels were then constructed around each position and summed together to create individual data tracks, which were then similarly quantile-normalized and combined as described above.

### 3.4.10 Putative origin dataset

To construct the putative origin dataset, first all 829 available ARS regions in the *S. cerevisiae* genome were obtained from OriDB (Nieduszynski et al. 2007). Next, the highest scoring ACS sequence was found within each ARS region. Every position on both the Watson and Crick strands was assigned a natural log-odds ratio consisting of the probability of a sequence matching the 33-bp ACS motif (Eaton et al. 2010) relative to the probability the sequence originated from a 4th-order Markov model of the background sequence. Only ARS regions containing a ratio above 4 were kept for further analysis, leaving 798 ARS regions after applying this filter. Finally, these origins were centered on the highest scoring ACS position and oriented by the T-rich ACS strand.

### 3.4.11 Origin efficiency determination

An Okazaki fragment sequencing library was obtained from NCBI Accession Number SRR566701 (McGuffee et al. 2013) and aligned to the *S. cerevisiae* genome using the paired-end alignment strategy detailed above. An origin efficiency metric
(OEM) (McGuffee et al. 2013) was derived around each putative ACS element by comparing the proportion of all Watson- and Crick-aligned Okazaki fragments ($wl$ and $cl$, respectively) within a 10 kb window left of the ACS to the proportion of all Watson- and Crick-aligned Okazaki fragments ($wr$ and $cr$, respectively) within a 10 kb window right of the ACS: $OEM = \frac{wl}{wl+cl} - \frac{wr}{wr+cr}$. All origins with $OEM < 0.05$ were assigned an origin efficiency of 0.

3.4.12 Replication timing determination

Each putative origin was assigned either “early” or “late” based on evidence of replication initiation in the presence of hydroxyurea. A 5-kb window surrounding the ACS of each putative replication origin was checked for an enrichment of BrdU reads. Reads from both BrdU datasets were combined and the average number of reads expected to fall into an individual 5-kb bin was determined by assuming a random distribution of sequencing reads across the entire genome. Under these conditions, the expected number of reads contained in any individual 5-kb bin can be modeled using a Poisson distribution. “Early” putative origins were defined as any 5-kb bin with $P < 1 \times 10^{-5}$ given this background distribution. All other origins were assigned “late”.

3.4.13 Individual gene locus plots

Two-dimensional plots were constructed similar to “V-plots” as described previously (Henikoff et al. 2011), with the x-axis representing genomic position and the y-axis corresponding to the fragment length of the sequencing read. Each sequencing read is individually plotted by first extracting the start ($st$) and end ($en$) coordinates along with the fragment length ($l$). The sequence read is then trimmed to half its fragment length, reassigning the start and end positions to $start = st + \frac{l}{4}$ and $end = en - \frac{l}{4}$, respectively. A blue rectangle is then filled with the following
coordinates: \((start, l - \frac{1}{2}), (start, l + \frac{1}{2}), (end, l - \frac{1}{2}), (end, l + \frac{1}{2})\). Reads are individually plotted so that denser blue regions correspond to the presence of more recovered reads. Aggregate two-dimensional plots (such as those shown in Fig. 3.2) are obtained by totaling the number of reads surrounding the ACS across multiple sites, with each aligned read oriented relative to the origin-specific ACS direction. Merge plots were constructed by assigning an RGB value to every square of the heatmap, where red represented the total signal from dataset one and green the signal from dataset two.

Cartoon schematics above each plot were computationally derived by finding peak positions in either the nucleosome (red) or small footprint (dark green) signal track. Nucleosome shading indicated the observed protein occupancy at a particular position, with darker red corresponding to a higher signal in the smoothed track.

3.4.14 Identifying replication origin footprints

Putative origins were first evaluated for the presence of an ORC-dependent footprint signal in the small fragment (25–120 bp) signal track. The total signal at positions -50–150 bp surrounding every ACS was obtained from both the G2 WT and \textit{orc1-161} conditions. Origins were classified as having an ORC-dependent signal if the total G2 WT signal in this region was above 100 and the fold-enrichment of G2 WT to G2 \textit{orc1-161} was above 1.5.

For those origins without an ORC-dependent footprint, a parallel approach was conducted, except this time comparing the G1 and G2 signals. Origins were classified as having a G1-Only Footprint signal if the total G1 WT signal in this region was above 100 and the fold-enrichment of G1 WT to G2 WT was above 1.5.
3.4.15  Cell-cycle–dependent nucleosome repositioning

G2 origin-flanking nucleosome positions were first determined for each origin in the G1 & G2 Footprint and G1-Only Footprint classes. Local maximum peaks in a 1-kb window surrounding each ACS were determined from the smoothed nucleosome signal track using the peakDetection function from the nucleR R package (Flores and Orozco 2011), requiring a minimum signal of 0.25 and ensuring that no other peak existed within 75 bp of the inferred nucleosome position. Then, each origin was assigned an upstream and/or downstream nucleosome position if the individual origin contained a nucleosome peak within 100 bp from the consensus upstream or downstream nucleosome positions (−90 bp and 148 bp relative to the ACS, respectively).

To find nucleosomes exhibiting cell-cycle–dependent dynamics, the G1 and G2 nucleosome signals were compared at each derived G2 nucleosome position. First, the G1 and G2 smoothed nucleosome signals were extracted ±100 bp around the G2 nucleosome position. Then, each smoothed nucleosome signal was converted to a probability distribution, representing the likelihood of a nucleosome existing at a particular position. 100 random nucleosome positions were drawn from the G1 and G2 nucleosome position probability distributions, and the median G1 and G2 nucleosome positions derived from their respective samples were compared. This sampling was repeated for 500 iterations. At each origin, if at least 90% of median G1 nucleosome positions were consistently greater or less than median G2 nucleosome positions, the flanking nucleosome was considered “dynamic”.

Nucleosome ratio heatmaps were constructed by first extracting the G1 WT, G1 cdc6-1, or G2 nucleosome density ±500 bp around each of the 398 origins in the G1 & G2 Footprint and G1-Only Footprint classes. This yielded three 398 × 1001 matrices from each dataset. To normalize for MNase digestion differences, the matrices were
quantile-normalized. Then, the log2 signal from G2 WT was subtracted from either the log2 G1 WT or log2 G1 *cdc6-1* samples, yielding a log2 ratio of nucleosome density at each position. Rows were ordered by the nucleosome shift classes as defined above. Genomic positions with higher ratios of G1 WT or G1 *cdc6-1* nucleosome density are colored red; likewise, higher ratios of G2 WT nucleosome density are colored green.

### 3.4.16 ORC and Mcm2-7 ChIP-seq analysis

The total ORC and Mcm2-7 ChIP-seq signal for each putative origin was determined by aggregating the smoothed signal track from $-400$ bp to $400$ bp relative to each ACS. A putative origin was assigned an ORC peak if any position contained a signal greater than $0.8$; likewise, an Mcm2-7 peak was assigned if any position contained a signal greater than $0.5$. The genomic coordinate with the highest signal was considered the peak position, allowing only one ORC and one Mcm2-7 peak to be assigned per origin.

389 G1 & G2 Footprint or G1-Only footprint origins (excluding 9 origins with insufficient ChIP-seq signal) were subdivided into 2 groups based on the Mcm2-7 ChIP-seq enrichment signal $\pm 100$ bp surrounding the upstream or downstream consensus G1 nucleosome positions ($-90$ bp and $162$ bp relative to the ACS, respectively). To better infer the peak signal position, strand-specific ORC and Mcm2-7 ChIP-seq reads were then plotted as a merged heatmap as described above, with reads aligning to the forward strand in the red channel and reads mapping to the reverse strand in the green channel.
Complete sequencing of genomes has revealed open reading frames that define protein-coding regions; however, the complete set of DNA sequences and binding factors responsible for regulating these genes largely remain unknown. Regulatory elements lack a defined array structure such as codons, confounding their precise identification by sequence alone. Further, access to potential regulatory regions may be obscured by nucleosomes, the principal factor responsible for compacting the genome. Thus, a fundamental understanding of where proteins are assembled along the genome will yield significant insights into the regulation of genomic processes such as transcription and replication.

Chromatin profiling first began with DNA footprinting experiments at individual promoters. These experiments utilized a nuclease such as DNase I or MNase to reveal protein-binding locations relative to a feature of interest. Evaluation of the chromatin environment in different growth conditions or on genetically-modified DNA templates uncovered key regulatory elements that impacted transcription. For example, footprinting experiments at the GAL1-10 locus revealed a dramatic nucleosome reorganization in the regulatory region during galactose-induced gene activation.
Similar chromatin dynamics were also found at the PHO5 locus (Almer et al. 1986; Boeger et al. 2003), demonstrating a conserved method of transcriptional regulation.

Even though these footprinting assays yielded significant insights, they were limited to individual promoters. With the advance of tiling arrays and then next-generation sequencing, these assays could be extended to profile chromatin genome-wide. Classical digestion of chromatin with MNase revealed nucleosome locations throughout the genome and a stereotypical nucleosome configuration surrounding transcription start sites (Sekinger et al. 2005; Yuan et al. 2005; Albert et al. 2007; Mavrich et al. 2008a,b). DNase I studies uncovered open chromatin regions where smaller DNA-binding proteins such as transcription factors could bind (Crawford et al. 2006; Sabo et al. 2006; Hesselberth et al. 2009). Each assay profiles a distinct subset of the chromatin environment, yet neither is sufficient to capture the entire chromatin architecture, nucleosomes and transcription factors, in one experiment.

Our MNase chromatin profiling method, by modifying the chromatin digestion and recovering all protected DNA fragments, provides an alternative assay that can profile the entire chromatin environment in one experiment. This provides a cost-effective method to simultaneously analyze nucleosome positions and smaller DNA-binding proteins. As the assay is agnostic, all proteins bound throughout the genome are captured. In contrast, chromatin immunoprecipitation can only profile one DNA-binding factor, and requires an available antibody to capture the protein of interest. Thus, profiling the chromatin environment under multiple conditions or across time points requires only one MNase experiment, as opposed to many ChIP assays, per condition. MNase epigenome mapping provides a comprehensive evaluation of the chromatin structure, and an independent validation of these previous genome-wide assays.

By simultaneously profiling both nucleosomes and DNA-binding proteins, the
genome can be partitioned into regions occupied by nucleosomes and those available for binding by smaller factors. This provides an opportunity to revisit previous binding annotations based on lower resolution ChIP-chip (Harbison et al. 2004; MacIsaac et al. 2006), which predicted binding sites without consideration of local nucleosome positioning. Utilizing information on the local chromatin environment, in conjunction with ChIP occupancy and evolutionary conservation, should improve predictions.

One potential downside of the MNase assay is deciphering the identity of proteins that are bound throughout the genome. Although searching for highly enriched motifs and using conservation information can gain insights into the likely binding candidates, this presents a complex problem when multiple factors are bound in close proximity. Deconvolving the occupancy profiles at these sites may require advanced algorithmic approaches. Experimentally, increasing the recovery of the smallest DNA fragments (20- to 50-bp fragments) would improve resolution, showing precisely which individual DNA bases are protected. Currently, this is a challenge because the Illumina library preparation steps and sequencing platform are optimized for ~150-bp DNA fragments. Further, longer MNase digestion times would be needed to allow the MNase enzyme ample opportunity to digest DNA fragments to the protein boundary. However, at these digestion times, MNase tends to outcompete transcription factors for access to the DNA sequences, evicting the smaller DNA-binding proteins from their in vivo locations. Improved crosslinking efficiency or better crosslinking agents may alleviate this problem.

Another potential limitation of the assay is the sequence bias exhibited by MNase upon DNA digestion. This bias results in the recovery of primarily DNA fragments that contain an adenine or thymine at the end position (Hörz and Altenburger 1981). As MNase digestion results in a much smaller percentage of recovered DNA fragments yielding a cytosine or guanine at the fragment end position, there is concern that the
MNase-derived chromatin profile is sequence-dependent (Chung et al. 2010). However, a recent nucleosome profiling experiment utilizing an independent chemical DNA cleavage method yielded nucleosome positions largely concurrent with MNase-derived locations (Brogaard et al. 2012). Further, methods to computationally correct for the MNase bias did not drastically alter nucleosome positions (Mavrich et al. 2008b; Tolstorukov et al. 2009; Gaffney et al. 2012). These results indicate that the MNase bias likely does not have a substantial effect on deriving nucleosome positions.

However, this MNase bias could impact the recovery of smaller DNA fragments. As the ultimate goal is to achieve base-pair resolution of transcription-factor binding, stretches of cytosines or guanines flanking a transcription-factor–binding site could adversely affect recovery of small DNA fragments in this region. To overcome this issue, MNase digestion could be combined with another exonuclease; in fact, more precise nucleosome positioning has been obtained by combining micrococcal nuclease with exonuclease III (Nikitina et al. 2013).

Comprehensive profiling of the chromatin environment through this MNase assay would have practical utility in higher eukaryotes such as Drosophila and human. In particular, the ENCODE project is attempting to analyze the chromatin state in hundreds of cell lines. As performing multiple chromatin immunoprecipitation experiments in many of these cell lines is not cost-effective, the MNase assay could be a useful tool to capture cell-type–specific differences in the chromatin state. However, the current limitation in expanding this protocol is the required sequencing library size. In yeast, ~20 × 10^6 reads are needed to comprehensively profile the chromatin. Scaling to the Drosophila and human genome sizes would thus require ~230 × 10^6 and ~5 × 10^9 reads, respectively. Although the number of reads required for Drosophila are practical given the current sequencing depth attainable with the Illumina HiSeq platform, the sequencing depth needed for human is currently cost prohibitive. Part of the reason for the high number of reads is to ensure sufficient recovery of small
fragments (20–120 bp); enhancing the proportion of these reads that are sequenced will help reduce the required library size.

As an example demonstrating the utility of the MNase assay, the chromatin environment surrounding DNA replication origins was profiled at different stages of the cell cycle. A high-resolution ORC footprint was observed at a subset of replication origins in G2, confirming previous results with lower-resolution ChIP-seq (Eaton et al. 2010). Although ORC in *S. cerevisiae* is bound at replication origins throughout the cell cycle, a subset of replication origins exhibited a more transient ORC interaction, only yielding a DNA footprint upon completion of pre-RC formation in G1. This suggests that ORC has different turnover kinetics at these replication origins, even though the ACS (the ORC binding motif) is indistinguishable from the replication origins harboring an ORC footprint in G2. As the MNase assay may not have the sensitivity to capture this transient ORC association, assessing ORC turnover by a competition ChIP approach (Lickwar et al. 2012) could yield more information about ORC-binding kinetics. These replication origins with a delayed footprint tended to be more inefficient, possibly handicapping the sequestration of limiting S-phase initiation factors like Cdc45.

Many replication origins exhibited chromatin dynamics upon pre-RC assembly in G1. Interestingly, each replication origin demonstrated a preferred orientation of pre-RC assembly and nucleosome movement. To accommodate the addition of Cdc6, a subset of replication origins shifted their upstream nucleosome away from the ACS. Similarly, a class of replication origins specifically moved their downstream nucleosome away from the replication origin. Both of these nucleosome shifts were dependent on Cdc6, yet it remains unknown if this nucleosome movement is mediated only by assembly of pre-RC components, or requires a chromatin remodeler to permit nucleosome sliding. Chromatin remodelers are enriched at particular nucleosome locations relative to transcription start and end sites (Yen et al. 2012). As
most replication origins are proximal to gene boundaries, it reasons that some of
these chromatin remodelers may play dual roles in affecting transcription and repli-
cation initiation. Analyzing the chromatin structure throughout the cell cycle in a
panel of chromatin remodeler deletion mutants may yield insights into which chro-
matin remodelers are necessary to induce these nucleosome dynamics. Chromatin
remodeling may also be required during pre-IC formation when additional proteins
are recruited to the replication origin. It will be interesting to see whether additional
nucleosome movement or eviction occurs at this stage, and whether recruitment of
specific chromatin remodelers impacts replication origin efficiency.

The directional nucleosome movement observed at replication origins may re-
fect restrictions in the local chromatin environment specifying the location of the
dynamic nucleosome. In particular, as predicted by the barrier model, transcription-
factor binding can limit the movement by adjacent nucleosomes. Using the MNase
assay to profile transcription-factor footprints surrounding replication origins may
reveal transcription-factor binding proximal to the nucleosomes that do not undergo
a shift during pre-RC formation. Further, this may explain why a subset of replica-
tion origins do not move either nucleosome upon pre-RC formation. Local binding of
certain transcription factors like Fkh1 and Fkh2 promote early-activating replication
origins (Knott et al. 2012), and additional transcription factors may enhance repli-
cation origin function. Alternatively, one origin-flanking nucleosome may occupy a
DNA sequence favorable for nucleosome formation, thus energetically predisposing
the complementary nucleosome for movement.

Like with the nucleosome movement and Cdc6-dependent footprint expansion,
Mcm2-7 loading is predominantly localized to one side at each replication origin.
However, the location of Mcm2-7 loading is independent of which nucleosome shifts,
suggesting additional mechanisms that direct Mcm2-7 loading. As Mcm2-7 loading
is in close proximity to a nucleosome, there may be a specific nucleosome modifica-
tion that attracts Mcm2-7. However, what function could localizing Mcm2-7 serve? Perhaps replication origins specify the Mcm2-7 location to the position where it is easiest to commence DNA unwinding. Or, a favorable chromatin environment at the Mcm2-7 loading position better facilitates recruitment of the pre-IC components. Profiling the chromatin environment at the beginning of S phase would reveal if pre-IC formation is concentrated at the Mcm2-7 loading position.

The chromatin environment clearly plays a significant role in replication origin location and function. Even though every replication origin assembles the same constituent proteins, assembly in the context of the local chromatin environment impacts origin efficiency and activation time, providing an additional layer of replication regulation. Although several chromatin features are correlated with replication origin function, such as ORC stability and nucleosome dynamics, neither completely determines origin efficiency or activation time. Ultimately, several aspects of the chromatin environment, such as nucleosome positioning, histone modifications, nearby transcription, and proximal DNA-binding proteins, collectively influence replication origin function. Future studies on the chromatin environment will likely reveal additional determinants of origin function.
Bibliography


Takara TJ, Bell SP. 2011. Multiple Cdt1 molecules act at each origin to load replication-competent Mcm2-7 helicases. *EMBO J.* **30**: 4885–4896.


Biography

Jason Alan Belsky was born on October 29, 1986, in White Plains, NY. He grew up in Yorktown Heights, NY, and received his high school diploma from The Masters School in Dobbs Ferry, NY, in 2004. He attended Cornell University for his undergraduate degree, receiving a BS in Biological Engineering in 2008. After a two-year post-baccalaureate IRTA fellowship at the National Institutes of Health in Bethesda, MD, he matriculated into the Computational Biology and Bioinformatics Program at Duke University in 2010.

Publications


