Nucleosome positioning in budding yeast

Posicionamiento de nucleosomas en *Saccharomyces cerevisiae*

Ozgen Deniz

UNIVERSITAT DE BARCELONA

FACULTAT DE BIOLOGIA

PROGRAMA DE DOCTORADO EN BIOMEDICINA

NUCLEOSOME POSITIONING IN BUDDING YEAST

Ozgen Deniz 2014

TESIS REALIZADA EN LABORATORIO EXPERIMENTAL DE BIOINFORMÁTICA

INSTITUT DE RECERCA BIOMÈDICA – BARCELONA

POSICIONAMIENTO DE NUCLEOSOMAS EN *SACCHAROMYCES CEREVISIAE*

Thesis submitted by - Memoria presentada por

Ozgen Deniz

To qualify for the Doctorate degree by - Para optar al grado de Doctor por la

University of Barcelona - Universitat de Barcelona

Thesis supervisor

El director de la Tesis

Modesto Orozco

To my family

Guzel aileme

ACKNOWLEDGEMENT

ACKNOWLEDGEMENT

First of all, I would like to thank my advisor Modesto Orozco, for giving me the opportunity to work in his laboratory, allowing me to pursue some of my ideas, for the helpful discussions, for his support and supervision. I am also grateful to him for always giving me the opportunities to improve my scientific background during conferences, short stay and collaborations.

I am indebted to Montse Soler for her support throughout my PhD and being always helpful, open and very patient (especially in my first year). Without her, EBL would not be like it is now. Montse, thank you for your time and understanding and helping me to improve my personality and scientific background.

I want to thank to Ferran Azorin for allowing me spend couple of months in his lab to learn chromatin techniques and Marti Aldea for scientific discussions.

I am grateful to all previous and present members of EBL for their help, creating a friendly working atmosphere and making hard times less painful. Thanks to Guillermo Suñe, Kathryn Collinet, Maica Lopez, Elisa Duran, Ricart Lluis, Chiara Catellazi and Rodrigo Arroyo for a nice and funny start and making EBL happen and Nahuai Badiola, Marianela Masin, Nuria Villegas, Victor Alcalde, Eva Capdevila, Isabelle Heath, Antonio Rodriguez, Jofre Tenorio, Montserrat Terrazas, Claudio Di Sanza, Cristina Battle for improving EBL even more and for nice discussions.

I have to mention my gratitude to the computational part of Orozco group for nice moments, parties and help for analysis. Especially to Oscar Flores, without him this thesis would not be possible. And to Federica Battistini for scientific help and being so much fun working with and partying with ☺

I am especially thankful to my friends that I met in Barcelona for sharing my good and bad moments, bearing me in my stressful moments, helping me during all my injuries and making my life funny and crazy. Thank you Zina, Michi, Ana, Peter, Sean, Luca, Radek, Jelena, Jordi, Thasso, Vanya, Mads, Sylwia for the perfect moments, crazy parties, explorative excursions, cozy dinners and being my friends… Thanks to my girls Marta, la pequeña, and Selma Beshosh for girly moments, listening to complains and funny lunch breaks and parties… Thanks to Andrey and Myk sharing everything with me

food, drinks, roof, bed-stories, fun, coffee breaks, my crying moments, a great friendship, …… Thanks to my previous flatmates Anna and Judith for the sistership that we had in Sant Elml 61…Thanks to my Turkish crowd : Emre, Besray, Baybars, Guzin, Kader and Gamze with whom I shared my most intimate moments. Emosum, sensiz havuzsuz, kisin denize girilmeden , dertlesmeden, Nilsi senden korumadan gecmezdi bu doktora… Kaderim yawrum, cidden basindan beri yanimda oldugun icin, hep beni dinledigin icin, super deli bi arkadas oldugun icin cok tesekkurler… Ve Baybars, her ne kadar bu doktorada yeteri kadar yanimda olamasan da, biliyorum hep olacaksin hayatimin bir yerlerinde…And lastly to Milica, she is almost in the Turkish crowd ☺ Thanks to my faunus girl, I am so glad you came out of the faunus and joined to me in this world… Without you I would not be able to make it, anything. Thanks for your support, making me laugh, making me cry, bearing difficult moments together with me and being another part of me. I wish us as many years together as our laughs and tears…

I also want to mention my far away friends that have been always my support : Eslaam, Berna, Lucia, Ceren,Gunes, Turan, Alper, …

And the Nils… Thank you Nilscim for the last 3,5 years that made me very happy. Thanks for taking me no matter how difficult I was and always believing in me. I would not survive my frustrations, breakdowns, desperations and will be lost without you being supportive, concerned and caring. Thank you being so special for me…

Lastly, I am very grateful to my parents, who are always supportive and my brother, with whom I am strongly bound by love. Canim ailem, bana hep guc verdiniz.. Hep yanimda olun… Sizi cok seviyorum…

CONTENTS

CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

ABSTRACT

The nucleosome is the fundamental structural unit of DNA compaction in eukaryotic cells and is formed by the wrapping of 147 bp double stranded DNA around a histone octamer. Nucleosome organization plays a major role in controlling DNA accessibility to regulatory proteins, hence affecting cellular processes such as transcription, DNA replication and repair (Khorasanizadeh 2004; Jiang & Pugh 2009).

Our study focuses on genome-wide nucleosome positioning in *S. cerevisiae* to explore nucleosome determinants and plasticity throughout the cell cycle and their interplay with gene expression based on cell mRNA abundance.

We pursued the contribution of DNA physical properties on nucleosome organization around key regulatory regions such as transcription start and terminaton sites (TSSs and TTSs) by analyzing genome-wide MNase-digestion profile of genomic DNA. We also implemented a systematic approach to standardize MNase-Seq experiments by minimizing the noise generated by extrinsic factors to enable an accurate analysis of the underlying principles of nucleosome positioning and dynamics. Moreover, we carried out a large-scale study of nucleosome plasticity throughout the cell cycle and its interplay with transcription based on a comparative analysis among nucleosome maps, gene expression data and MNase sensitivity assays. We then focused on nucleosome organization around DNA replication origins (ORIs) and its possible effect on origin activation. Finally, we sought to characterize centromeric nucleosome composition and its oscillation along cell cycle.

During the course of these studies, we found that key regulatory regions such as the nucleosome free regions (NFRs) contain unusual physical properties that are intrinsic to genomic DNA. We further demonstrated that DNA physical properties and transcription factors act synergistically to define NFRs, especially in genes with an open promoter structure. Once NFR is defined, the nucleosome positioning around TSSs can be predicted by a simple statistical model, supporting the energy barrier model for nucleosome positioning. However, we also observed that nucleosomes are quite dynamic at distal 5' NFRs and do have distinct regulatory mechanisms.

Our comparative analysis of nucleosome organization along cell cycle revealed that chromatin exhibits a distinct configuration due to DNA replication-

dependent organization at S phase, showing higher sensitivity to MNase and displaying fuzzier nucleosomes along the genome. Moreover, we observed different features at M phase, where chromatin compaction is the highest and displays a slightly different pattern than in G1 and G2 phases. Interestingly, these changes in chromatin organization are sudden and acute and only affect some regions of the genome, whereas the majority of genes present conserved nucleosome patterns along cell cycle. Our individual gene analysis disclosed that the largest changes in nucleosome architecture take place in cell cycle-dependent genes, indicating the interplay between chromatin structure and transcription.

The detailed analysis around replication origins shows that they display slightly wider NFRs at G1 phase due to pre-Replication complex binding. Once the complex disassociates, nucleosomes partially occupy NFRs, but up to a certain extent due to constitutive binding of ORC. We provided evidence that early firing origins tend to have more ordered nucleosome organization than late firing origins.

Finally we illustrated that centromeric nucleosomes display a perfect positioning, confirming their strong centromeric sequence-dependent recruitment to DNA. The characterization of histone composition under physiological cell conditions suggested that the octameric nucleosome assembly model is favored in centromeres. Yet, our analysis along cell cycle showed centromeric nucleosome dynamics, proposing that its composition might oscillate along cell cycle.

Taken together, our accurate study provides a dynamic picture of nucleosome positioning and its determinants; new insights into cell cycle-dependent chromatin organization on key regulatory regions and its interplay with gene expression; and adds a new dimension to the characterization of centromeric nucleosomes.

ABBREVIATIONS

ABBREVIATIONS

2x	Paired End Sequencing
A	Adenine
APC/C	Anaphase-Promoting Complex/Cyclosome
3C	Chromosome Confirmation Capture
ACS	ARS Consensus Sequence
AFM	Atomic Force Microscopy
APS	Ammonium Persulfate
ARS	Autonomously Replicating Sequence
As	Asynchronous
ASF-1	Anti-Silencing Factor 1
ATP	Adenosintriphosphate
bp	Base Pair
BSA	Bovine Serum Albumin
C	Closed (NFR)
C	Cytosine
CAF-1	Chromatin Assembly Factor 1
CATD	CENP-A Targeting Domain1
CBF	Centromere Binding Factor
CDB	Coverage Difference Per Base
CDE	Centromere DNA Element
CDI	CDK Inhibitor
CDK	Cyclin Dependent Kinase
CEN	Centromeric
CENH3	Centromeric H3
CENP-A	Centromere Protein A
CHD	Chromodomain-Helicase-DNA-binding
CHIP	Chromatin Immunoprecipitation
CID	Centromere Identifier
Clb	Cyclin B
Cse4	Chromosome Segregation 4
CTD	Carboxyl Terminal Domain
C-terminal	Carboxyl Terminal
Da	Dalton
DMSO	Dimethyl Sulfoxide
EDTA	Ethylenediaminetetraacetate-disodium
EGFP	Enhanced Green Fluorescent Protein
EM	Electron Microscopy
EtBr	Ethidiumbromide
F	Fuzzy
FACS	Fluorescence Activated Cell Sorting
FCS	Fluorescence Correlation Spectroscopy
FKH	Forkhead Transcript. Factors
FRET	Fluorescence Resonance Energy Transfer
G	Guanine
G1	Gap 1
G2	Gap 2
GA	Genome Analyzer
GO	Gene Ontology
H	Histone
h	Hour
HDAC	Histone Deacetylase
Ino80	Inositol Requiring
Isw1/2	Imitation Switch
LR	Low Coverage Region
M	Missing / Mitosis
MCM	Mini-chromosome Maintenance
MD	Molecular Dynamics
MNase	Micrococcal Nuclease

CLR	Common Low Regions
NFR	Nucleosome Free Region
NPS	Nucleosome Positioning Signal
N-terminal	Amino Terminal
O	Open (NFR)
OD	Optical Density
RPA	Replication Protein A
rpm	Revoltations Per Minute
r.p.m.	Reads Per Million
RSC	Remodel The Structure Of Chromatin
RT	Room Temperature
S	Synthesis
S. cerevisiae	*Saccharomyces Cerevisiae*
S. pombe	*Schizosaccharomyces Pombe*
SD	Standard Deviation
SDS	Sodiumdodecylsulfate
SDS-PAGE	SDS-Polyacrylamide Gel Electrophoresis
Seq	Sequencing
NAP-1	Nucleosome Assembly Protein 1
Snf	Sucrose Non-fermenting Protein
SWI/SNF	Switch/Sucrose Non-fermenting
Swr1	Swi2/Snf2-related 1
T	Thymine
TBE	Be Tris-Borate-Edta
Temed	Tetramethylethylenediamin
TF	Transcription Factor
TFBS	Transcription Binding Site
TSS	Transcriptional Start Site
TTS	Transcription Termination Site
Un	Under-Digested
UV	Ultraviolet
W	Well-positioned
w/v	Weight Per Volume
YPD	Y*east* Peptone Dextrose

INTRODUCTION

1 Chromatin Structure

1.1 Overview

The most important function of the DNA is to carry the genetic information, encoded by the genetic code that specifies the proteins to be synthesized. In eukaryotic cells, most of the DNA is enclosed in the nucleus, whereas some of the DNA is stored in mitochondria or chloroplasts (Stocking & Gifford 1959). Considering that the linear length of the DNA could reach several meters, DNA has to be packed tightly in a way to fit into the nucleus. To achieve this firm compaction, DNA is organized into the fibrous network called "chromatin". The term "chromatin" was first characterized by Walther Flemming (Flemming 1882), who defined it as "stainable material of the nucleus". Besides its role in condensing DNA within nucleus, chromatin also functions to control the accessibility of DNA to DNA-binding proteins. Hence, chromatin regulates a variety of DNA metabolism processes like transcription, DNA replication, DNA recombination and DNA repair (Khorasanizadeh 2004; Jiang & Pugh 2009; Felsenfeld 1992). Depending on the transcription status or cell cycle stage, chromatin adopts either a more loosened or condensed conformation to regulate the accessibility of DNA. Chromatin structure is then crucial for gene regulation and its level of compaction is tightly regulated.

1.2 Nucleosome: Chromatin Primary Structure

In eukaryotes, the fundamental structural repeating unit of chromatin is the nucleosome (Kornberg 1974). Early studies demonstrated that in many eukaryotes the digestion of chromatin by enzymatic nucleases results in a protected and conserved DNA fragment of 146 ±2 bp (Van Holde n.d.). Later in 1997, the atomic structure of nucleosome was determined by X-ray crystallography (Luger et al. 1997) (**Figure 1A**). The nucleosome is composed of 146 bp of DNA wrapped in 1.67 left-handed superhelical turns around an octamer of core histone (H) proteins; dimer of each H2A, H2B, H3 and H4 (Luger et al. 1997). To achieve tight wrapping around the histone octamer, nucleosomal DNA is sharply bent at every ~10 bp, when the minor groove of DNA faces inwards toward the histone core, and 5 bp away when the minor groove faces outwards (Segal et al. 2006) (**Figure 1B)**. All of the core histones have an extended histone-fold domain at their carboxyl terminal domain

(CTD), which consists of a long central α-helix flanked by a loop and a short α-helix. This domain is responsible for DNA-protein, as well as protein-protein intermolecular interactions, which assist the formation of H2A/H2B and H3/H4 dimers. Furthermore, histones have charged tails at the amino terminal (N-terminal) end, which are sites subjected to post-translational modifications

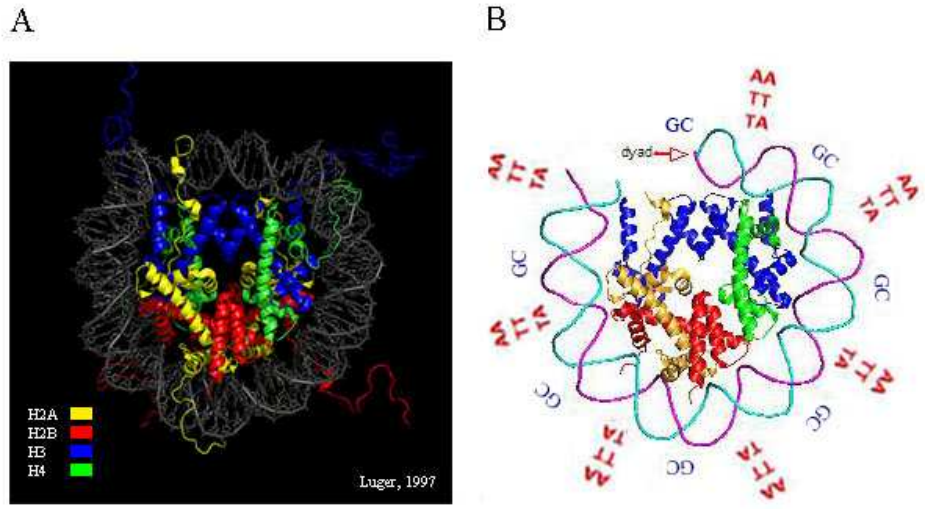


Figure 1: The nucleosome core particle. A) Crystal structure of the nucleosome core particle (Luger et al. 1997). The nucleosome is composed of 146 bp of DNA wrapped in 1.67 left-handed turns around an octamer of core histone proteins (dimer of each H2A, H2B, H3 and H4). B) nucleosomal DNA is sharply bent and this bending occurs at every ~10 bp when the minor groove of DNA faces inwards towards the histone core and 5 bp away the minor groove faces outwards (adapted from (Tolkunov & Morozov 2010))

(Wolffe 1998). In most organisms, in addition to core histones, there is a fifth histone called linker histone (H1 or its isoform H5) that associates with DNA at the entry or exit of the nucleosome (Szerlong & Hansen 2010; Wolffe 1998). Linker histones are the most loosely bound histones to DNA. They play a crucial role in the folding of nucleosomes into a higher order structure (Woodcock & Ghosh 2010; Caterino & Hayes 2010).

Since the first crystallographic studies, a variety of structural analyzes provided detailed insights into the nucleosome structure. The structure analysis from distinct DNA and histone sequences exhibited that overall nucleosome structure is quite conserved regardless of the variations in DNA and protein sources (Luger et al. 1997; Robert K Suto et al. 2000; White et al. 2001; Davey et al. 2002; Bao et al. 2006; Tsunaka et al. 2005; Luger et al. 2012). Accordingly,

the canonical nucleosome structure has a pseudo-two-fold rotational symmetry (Luger & Richmond 1998), where DNA is divided from the axis into 73 and 72 bp halves with a central base pair (bp) on dyad (Luger et al. 1997). The superhelix, with an average diameter of 41.8 Å, is not uniformly bent, however in certain regions the curvatures are sharper (Luger et al. 1997). The protein octamer is divided into four histone fold dimers defined by H3-H4 and H2A-H2B dimers, where first $(H3\text{-}H4)_2$ tetramer binds to 60 bp long central nucleosomal DNA and then H2A-H2B dimers are organized toward both ends of the DNA (Luger et al. 1997). Each histone dimer has three distinct DNA binding sites, thus twelve in the octamer, formed by the contact of individual histones. In the octamer, 121 bp of DNA is bound to these 12 sites, while the rest of the DNA is bound to H3 histone fold extension resulting in more than 120 direct interactions between DNA and histones (Luger et al. 1997; Luger & Richmond 1998). These DNA-histone interactions take place every 10 bp and are mediated by salt bridges, hydrogen bonds and hydrophobic contacts with DNA backbone (Battistini et al. 2010) that, together with protein-protein interactions within the histone octamer, stabilize the nucleosome structure (Luger et al. 2012).

Even though the octasome structure has been revealed in details and is preferentially used to entitle the canonical nucleosome, nucleosomes should not be considered a unique unit, but a dynamic family of particles. Consistently, several non-canonical nucleosome structures have been reported (Zlatanova et al. 2009). The distinct structures of nucleosomes have different compositions of histones and different configurations, like tetrasome, hexasome, hemisome and lexasome. Even though their biological significance is not clear yet, it is likely that they may exist *in vivo* as the intermediates of assembly and disassembly processes (Luger et al. 2012).

In the nucleus, the nucleosomes are deposited into long arrays, where each nucleosome core particle is separated from each other by non-nucleosomal DNA, called linker DNA. The length of the linker DNA varies between 10 and 90 bp among different species, indicating that ~70-90 % of genomic DNA is wrapped around nucleosomes. At low salt concentration and without linker histones, nucleosomal arrays form an extended primary structure of 10 nm diameter. This primary structure of chromatin is referred to as “beads on a string”, due to its appearance by electron microscopy, where nucleosomes resemble beads and linker DNA a string (Olins & Olins 1974). The main

functions of the 10 nm chromatin fiber are the neutralization of negative charges on the DNA backbone by the histone positive charges and to provide the first level of DNA compaction to fit into the nucleus (Fussner et al. 2011). However, the 10 nm primary structure only provides 6-7 fold compaction (Kornberg 1974), suggesting that further compaction is required for chromatin organization in nucleus (see in section 1.3).

1.3 Chromatin Higher Order Structure

Klug *et al.* first proposed that chromatin fiber forms a higher ordered structure, which is referred as the secondary structure of chromatin (Finch & Klug 1976). This second level of packing involves the coiling of nucleosomes into a defined 30 nm fiber, which is stabilized by linker histones and provides another 60-fold compaction. Despite the intense amount of studies since the first proposal, the precise organization and structure of 30 nm fiber is still under debate. Early electron microscopy (EM), atomic force microscopy (AFM) and X-ray diffraction studies have led to two competing models: the solenoid and the zigzag models (Woodcock & Ghosh 2010; Robinson et al. 2006) (**Figure 2**). In the solenoid model, adjacent nucleosomes are connected by a bent linker DNA and follow a superhelical path resulting in a one-start (Widom & Klug 1985) (**Figure 2A**). Conversely, in the zigzag model, the adjacent nucleosomes are connected by a straight linker DNA in a zigzag shape resulting in two-start helix (Woodcock & Ghosh 2010; Williams et al. 1986) (**Figure 2B**). Even though tetranucleosome crystal structures and disulfide crosslinking studies strengthen the probability of the zigzag pattern (Schalch et al. 2005; Dorigo et al. 2004), a recent molecular tweezers experiment points to the solenoid model as the structure of the 30 nm fiber (Kruithof et al. 2009). Yet, most of our current knowledge about the secondary structure comes from *in vitro* data or from isolated chromatin (Schalch et al. 2005; Woodcock & Ghosh 2010; Robinson et al. 2006).

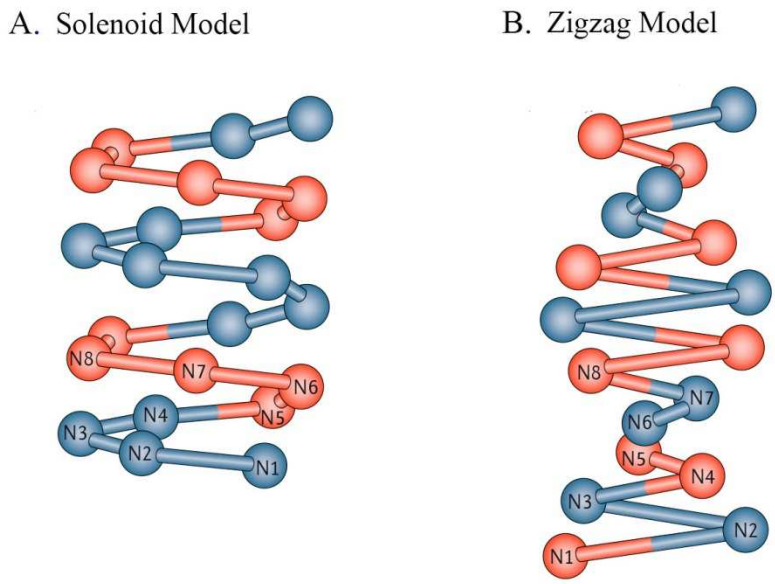


Figure 2: Two models for secondary structure of chromatin. A) Solenoid model is characterized by interactions between adjacent nucleosomes (n, n+1), which follow a superhelical path resulting in a one star helix structure. B) Zigzag model involves interaction between alternate nucleosomes (n, n+2), where the nucleosomes are arranged in a zigzag shape two star helix structure. The adjacent nucleosomes are numbered from N1 to N8. (Adapted from (Luger et al. 2012)).

The first *in vivo* evidence of the 30 nm fiber came from starfish sperm chromatin. Using both conventional EM and electron spectroscopic imaging, it was shown that the 10 nm primary structure is folded into a 30 nm width fiber (Horowitz et al. 1994; Bazett-Jones 1992). Moreover, the presence of 30 nm structure was also shown in chicken erythrocytes and mouse photoreceptors (Scheffer et al. 2011; Kizilyaprak et al. 2010). On the other hand, cryo-EM visualization of mitotic HeLa cells failed to demonstrate the presence of 30 nm fiber structure (Joti et al. 2012), although the study might suffer from structural artifacts arising from mitotic chromosome isolation. Yet, correlative electron spectroscopic imaging studies on induced pluripotent stem cells also demonstrated that heterochromatin is composed of 10 nm rather than 30 nm fibers (Fussner et al. 2011). Later, atomic force microscopy, cryo-EM and EM studies revealed heterogeneity on the nucleosome packing (Grigoryev et al. 1999; Bednar et al. 1998) where certain regions exhibit higher compaction while others show more loosened organizations. They further claimed that the level of compaction may depend on linker histones, linker DNA length variations and histone modifications.

Unlike higher eukaryotes, the budding yeast *Saccharomyces cerevisiae* has unique chromatin properties. *S. cerevisiae* does not have a functional linker histone and its linker DNA is relatively shorter compared to other higher organisms, which might both affect the organization of the higher order chromatin structure. Various studies in *S. cerevisiae* have contributed to the detailed higher order chromatin structure. Using EM, light scattering and X-ray diffraction, Lowary and Widom demonstrated that yeast chromatin undergoes folding into a 30 nm structure (Lowary & Widom 1989). Furthermore, deletion of the putative linker histone, Hho1p, in yeast has shown to modify the higher order chromatin organization (Georgieva et al. 2012). On the contrary, a study using chromosome confirmation capture (3C) has reported that yeast chromatin does not form a compact, but rather an extended fiber (Dekker 2008). This inconsistency between studies might be due to different histone modifications, variations in linker DNA length or limitations of the techniques used. Further studies are thus needed to confirm the secondary structure of the chromatin in yeast.

Even though the 30 nm structure is still elusive, some studies have suggested even further compaction into the nuclear scaffold or nuclear matrix, consisting of RNA and proteins (Hancock 2000; Fisher & Merkenschlager 2002). They claim that chromatin is organized into distinct domains. However, the existence of such a compact structure is still controversial (Bode et al. 2003).

Collectively, these results suggest that the 30 nm structure might exist although it would not be the predominant organization in chromatin and should be limited to a particular cell type or local condition.

2 Genome-wide Nucleosome Positioning in *S. cerevisiae*

Nucleosome positioning is defined as the location of nucleosomes with respect to the genomic sequences. High-resolution nucleosome maps reveal that nucleosomes can occupy multiple and overlapping positions, suggesting that nucleosomes are dynamic and might position at several places in the genome (Tanaka et al. 1996; Fragoso et al. 1995). Nucleosome positioning modulates eukaryotic gene regulation, since it controls the accessibility of DNA to the regulatory proteins. Thus, it influences many cellular functions like transcription, DNA replication and DNA repair, as mentioned before (Luger et

al. 1997; Felsenfeld 1992). Therefore, determining the exact positioning of nucleosome is key to understand the underlying fundamentals of complex cellular processes.

In recent years, many techniques have been developed for mapping nucleosomes *in vivo*. Principally, these techniques encompass isolation of nucleosomal DNA from chromatin followed by the subsequent identification of those nucleosomes within the genome. Isolation of nucleosomes usually involves chromatin digestion by nucleases. In particular, Micrococcal nuclease (MNase) is the most widely used nuclease, with an endo-exonuclease activity that cleaves both single- and double-stranded DNA and then trims the DNA from the exposed ends (Dingwall et al. 1981). MNase predominantly cuts at sites centered on A/T- containing dinucleotides (Dingwall et al. 1981). In chromatin, MNase cleaves the unprotected linker DNA, while it leaves protected nucleosome fragments intact. Once the protected nucleosomal fragments are isolated, they can be mapped along the genome by various techniques. Initial positioning studies were only able to determine the exact positions of a few hundred nucleosomes, due to technical limitations (Lee et al. 2004; Sekinger et al. 2005). Nevertheless, these ground-breaking studies established for the first time that the intergenic regions are usually depleted of nucleosomes (Lee et al. 2004; Sekinger et al. 2005). Later, a study by Yuan *et al*. provided the first high-resolution genome-wide nucleosome map, generated using a tiling microarray approach (Yuan et al. 2005). Yuan *et al*. defined a stereotyped chromatin organization around the gene coding start sites, consisting of a nucleosome free region (NFR) flanked by nucleosomes (Yuan et al. 2005). NFR was shown to be the preferred binding site for the majority of functional transcription factors (TFs), although some are able to bind upstream the NFR despite of nucleosome occupation (Rando & Chang 2009). Furthermore, NFRs seem to contain rigid poly (dA:dT) tracts that have poor nucleosome affinity (Yuan et al. 2005). In 2007, the first high-resolution genome-wide nucleosome map was generated by Lee et al (Lee et al. 2007), where 70 000 positioned nucleosomes, covering 81% of the yeast genome, were identified. This study further confirmed the canonical pattern of nucleosome positioning around the coding start region. Moreover, they demonstrated an inverse correlation between gene expression and nucleosome abundance on promoter regions; promoters of highly expressed genes are less occupied by nucleosomes, while lowly expressed genes usually contain more nucleosome-occupied promoters. The last breakthrough in the

field has been the development of massive parallel sequencing of nucleosomal DNA, providing single base-pair resolution nucleosome maps (Albert et al. 2007; Mavrich et al. 2008). The first study using this technology was the genome-wide positioning of nucleosomes containing the histone variant H2A.Z. The presence of 5' end NFR was confirmed at 95% of the genes as well as the presence of 3' end NFR and phased nucleosomes at the gene ends (Mavrich et al. 2008). To date, many nucleosome maps have been reported at a single base-pair resolution in yeast, worms, flies and humans(Albert et al. 2007; Mavrich et al. 2008; Schones et al. 2008; Shivaswamy et al. 2008).

Collectively, different genome-wide nucleosome maps obtained from distinct organisms have demonstrated that most of the nucleosomes are "well-positioned" or "phased", meaning that the nucleosome is positioned exactly at the same genome location within a cell population. Particularly in *S. cerevisiae,* about 80% of the nucleosomes are well-positioned and separated from each other by around 18 bp long linker DNA (Yuan et al. 2005; Mavrich et al. 2008; Lee et al. 2007). On the other hand, some nucleosomes, especially in the intergenic regions, are less localized and are referred to as "fuzzy" nucleosomes, which may occupy different locations on the genome within a population (**Figure 3A**). Moreover, several studies, following the original work of Yuan *et al.*, have revealed that nucleosome positioning across genome is not random but follows a canonical nucleosome pattern around transcriptional start sites (TSSs) that consists of a NFR flanked by two well-positioned nucleosomes (**Figure 3B**). The nucleosome immediately upstream the NFR is referred to as "-1 nucleosome", whereas the nucleosome downstream NFR is referred to as "+1 nucleosome" (Mavrich et al. 2008). The -1 nucleosome is subject to many changes, i.e. histone acetylation, repositioning and eviction, which usually affect its stability. On the other hand, the +1 nucleosome has the tightest positioning and often contains the histone variant H2A.Z (Albert et al. 2007; Mavrich et al. 2008). Moreover, +1 nucleosomes usually have the highest occupancy, which is defined as the nucleosome density or histone abundance in a given population.

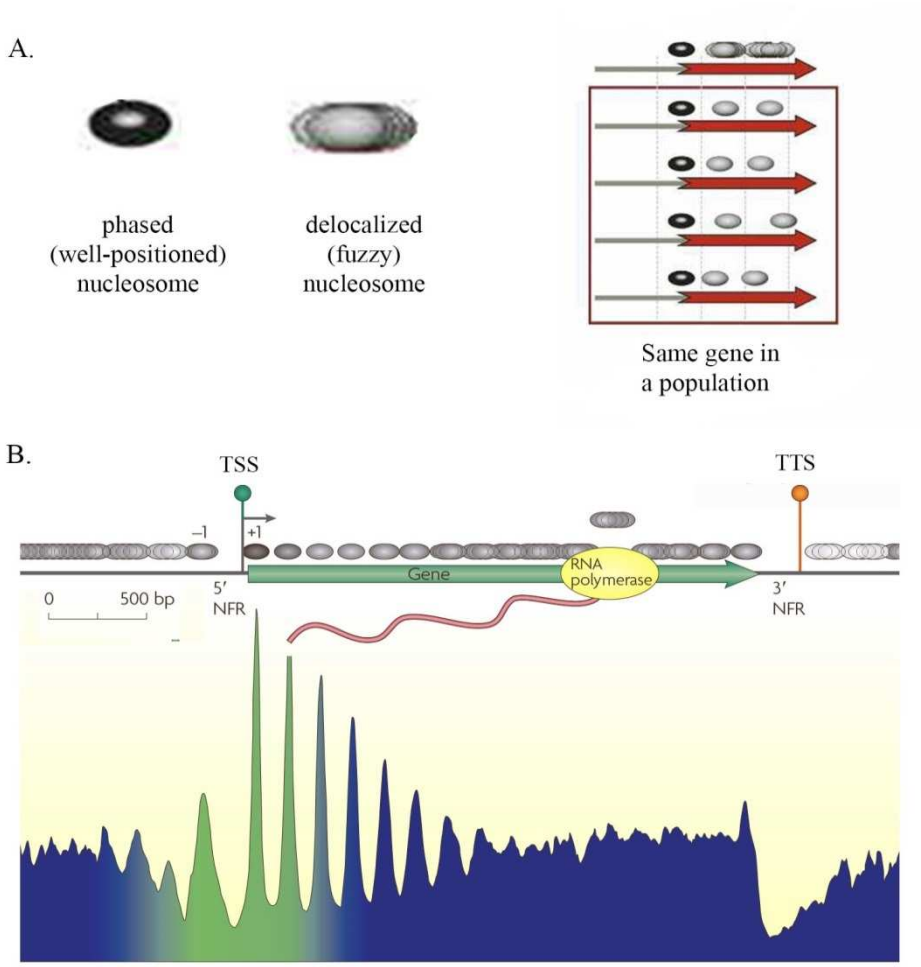


Figure 3: Nucleosome phasing and nucleosome organization. A) Phasing information of the nucleosomes. In a population nucleosomes are either positioned exactly at the same location in the genome, phased (well-positioned), or are less localized and spread out from the position, delocalized (fuzzy). B) Nucleosome pattern along the yeast genes with relative to transcriptional start site (TSS) and transcriptional termination site (TTS). (adapted from (Jiang & Pugh 2009; Mavrich et al. 2008)).

In coding regions, nucleosomes are arranged in a regularly spaced array starting at the +1 nucleosome. The phasing of nucleosomes decreases from the 5' gene end toward the 3' end (Albert et al. 2007; Mavrich et al. 2008) and terminates with a 3' end NFR, coinciding with the transcriptional termination site (TTS) (Jiang & Pugh 2009). This regular nucleosome array has brought the question of how this organization is established. One view suggests that the adjacent nucleosomes are individually and independently positioned (Jiang & Pugh 2009). An alternative view, called "statistical positioning" or "barrier"

model, was first proposed by Kornberg and Stryer (Kornberg & Stryer 1988) and later supported by various studies (Mavrich et al. 2008; Rando & Ahmad 2007; Ioshikhes et al. 2006). According to that model a barrier favors the formation of a well-positioned nucleosome, which in turn forces the positioning of neighboring nucleosomes (Mavrich et al. 2008). Therefore, a single barrier can position many nucleosomes without the need of an individual positioning signal for each nucleosome (Sadeh & Allis 2011). However what sets the barrier is still under debate. Some groups have suggested the transcription factors as the barriers (Kornberg & Stryer 1988), whereas Yuan *et al.* have proposed the poly(dA:dT) tracts and Mavrich *et al*. have suggested the +1 nucleosomes (Yuan et al. 2005; Mavrich et al. 2008). Yet, since there is evidence for every alternative, they may not be mutually exclusive.

3 Determinants of Nucleosome Positioning

Despite the tremendous amount of studies on nucleosome positioning, exactly what determines nucleosome positioning is still under debate. Two controversial factors have been suggested as the determinants of nucleosome positioning: *cis*-acting and trans-acting factors. Albeit intense research efforts have revealed that nucleosome positioning is not solely determined by one factor, but rather by the combined effects of many factors like DNA sequence, chromatin remodelers, posttranslational histone modifications, histone variants, TFs or RNA polymerase II (van Bakel et al. 2013; Struhl & Segal 2013).

3.1 *Cis*-acting Factors

Genome-wide maps and available nucleosome structures have allowed us to determine the effect of local factors, also called *cis*-factors. Basically, *cis*-factors are determined in terms of DNA sequence and its derived parameters, i.e. physical properties of the local DNA helix (Jansen & Verstrepen 2011).

Although histone octamers are practically able to bind to any DNA sequence, selectivity studies have exhibited that some sequences contain 10^3-fold higher affinity than random ones, demonstrating considerable DNA sequence specificity of the histone octamer (Thåström et al. 2004). However, unlike the other DNA binding factors, histones only have few base-specific contacts with

the DNA and little sequence preferences. Therefore, the specificity of nucleosome formation largely reflects the overall ability of the nucleosomal DNA to bend around the histone octamer (Drew & Travers 1985). This DNA bending occurs at every 10-11 bp helical repeat and introduces a deformation in the double-helical structure. In view of that, DNA sequences with higher bendability should be in contact with the histones for the optimal nucleosome formation, and similarly, stiffer DNA sequences should be mostly located at NFRs or linker DNA.

Physical models are often based on the core nucleosome crystal and NMR structures. These structures have revealed that DNA bends anisotropically around nucleosome, where sharp bends or mini-kinks are responsible for the tight wrapping of the double helix (Tolstorukov et al. 2007). The location of these bends depends on the helical geometry of the dinucleotide step in the double-stranded DNA, which can be characterized by six degrees of freedom; three translational (rise, shift, slide) and three rotational (twist, roll, tilt) (Trifonov 2010) (**Figure 4**). The 6 degrees of freedom are usually employed in many studies to derive empirical models of DNA elastic energy for predictive nucleosome positioning (Olson et al. 1998). Tolstorukov *et al.* first demonstrated that DNA double-helix folding is principally described by the roll and slide parameters (Tolstorukov et al. 2007), where positive slide and negative roll values accompany DNA bending into the minor groove whereas negative slide and positive roll values account for DNA bending into major groove of DNA (Tolstorukov et al. 2007). These “hot-spot” kinks occur at the most flexible dinucleotide base pairs, such as d(CA/TG) and d(TA/TA) dimers. However, another approach from Morozov suggested a more sensitive model to variations in the slide parameter (Morozov et al. 2009). On the contrary, Vaillant et al. demonstrated that only the roll parameter is actually sequence-dependent and contributes to nucleosomal DNA geometry (Vaillant et al. 2007). Indeed, DNA bending is dominated by the roll parameter that, together with an increased flexibility at kinks, determines the DNA sequence selectivity for wrapping around the nucleosome (Battistini et al. 2010; Battistini et al. 2012).

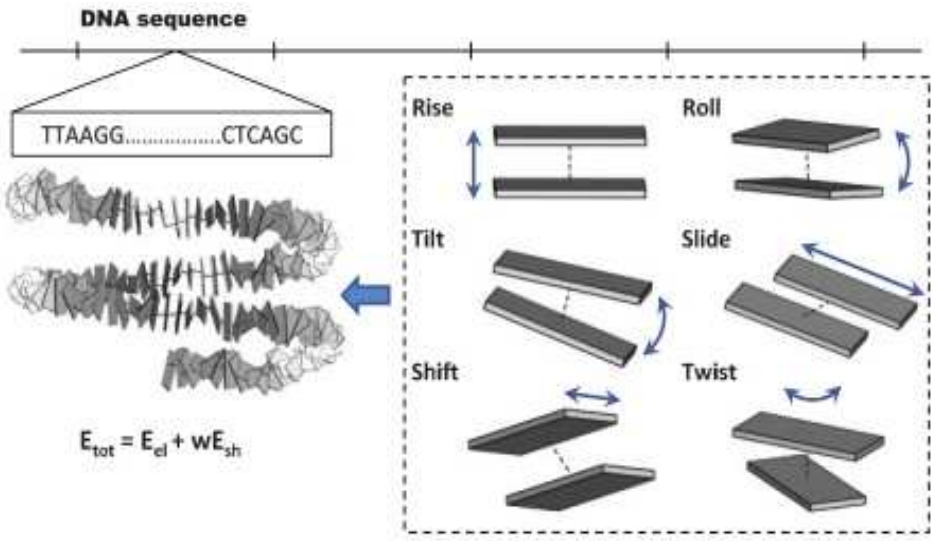


Figure 4: DNA mechanics model of histone–DNA interactions. Dinucleotide step in the double-stranded DNA is characterized by six degrees of freedom; three translational (rise, shift, slide) and three rotational (twist, roll, tilt; on the left). DNA base pairs are shown as rectangular blocks. The minimized nucleosome energy, which is the sum of elastic energy EeI and resistant energy Esh is represented on the right side (adapted from (Morozov et al. 2009)).

On the other hand, non-physical bioinformatics models determine the *cis*-factors by mining high-throughput nucleosome positioning data. In this approach, nucleosome-favoring and nucleosome-excluding sequence motifs are characterized and then used as training sets to predict common sequence patterns (Jansen & Verstrepen 2011). Ioshikhes and his colleagues analyzed dinucleotide distributions based on a collection of ~200 well-positioned nucleosomes from budding yeast, and disclosed a nucleosome positioning signal (NPS) of AA and TT with a periodicity of 10 bp (Ioshikhes et al. 2006). They found a particular high correlation of NPSs especially with +1 and -1 nucleosomes of TATA-less promoters, whereas a strong anti-correlation was observed at NFRs. On the contrary, TATA-box promoters exhibited weaker NPSs around TSSs and are actually mostly regulated by chromatin remodelers. Segal et al. used a different set of 199 mononucleosomal DNA sequences to parameterize another model (Segal et al. 2006), which exhibited a sequence motif of ~10 bp periodic AA/TT/TA dinucleotides oscillating in phase with each other and out of phase with ~10 bp periodic GC dinucleotides. Using a probabilistic model of dinucleotide frequencies, they compared the predicted nucleosome positioning with *in vivo* data. They were able to predict nearly 50% of nucleosomes within 35 bp of their *in vivo* positions. Even though

Ioshikhes and Segal's models seem to have predictive power when compared against *in vivo* data, Zhang *et al.* claimed that these sequence-based nucleosome signals actually account for 15-20% of nucleosomes in the genome (Zhang et al. 2009). Unlike the dinucleotide distribution approach, Peckham et al. has checked k-mer distributions (k=1 to 6 nucleotides) in 1000 highest and lowest scored nucleosomes from Yuan *et al.* data set. Their findings demonstrate that AT rich k-mers are in nucleosome-excluding sequences, whereas GC rich k-mers are in nucleosome-favoring sequences (Peckham et al. 2007). However, the authors estimate that 22-25% of nucleosome positions, albeit better than Segal and Ioshikhes, were due to DNA sequence signals.

The importance of DNA sequence on chromatin structure has been further tested by *in vitro* nucleosome reconstitution of yeast genomic DNA with histone octamers from chicken erythrocytes (Kaplan et al. 2008). Additionally, an *in vivo* nucleosome map has also been generated for a comparison. Consistent with the previous studies, *in vivo* and *in vitro* maps display a clear ~10 bp dinucleotide periodicity, although the dynamic range of the periodicity is greater *in vitro* (**Figure 5**). Accordingly, ~10 bp periodic AA/AT/TA/TT dinucleotides oscillate in phase with each other and out of phase with ~10 bp periodic CC/CG/GC/GG dinucleotides (**Figure 5A,B**). Analysis of 5-mer sequences has demonstrated that AAAA has the lowest nucleosome occupancy both *in vivo* and *in vitro*, in agreement with the abundance of poly (dA:dT) sequences at NFRs. The presence of a 5´ end NFR both *in vivo* and *in vitro*, albeit higher depletion degree in the *in vivo* map, indicates that nucleosome depletion around TSSs is partially encoded by intrinsic DNA properties, with a probable contribution of regulatory proteins. Furthermore, across all base pairs, the correlation of *in vitro* and *in vivo* maps in terms of nucleosome occupancy seems to be as high as 0.74. The correlation was higher at intergenic regions and TTSs, but much lower at promoter regions. Finally, even though the *in vitro* map shows an increased signal at -1 and +1 sites, it failed to accurately localize well-positioned -1 and +1 nucleosomes.

All of these studies thus demonstrate that only a small subset of nucleosomes is determined exclusively by DNA sequence *in vivo*. Conversely, NFRs can indeed be mostly predicted based on DNA sequence. Therefore, DNA sequence-based analyzes are commonly used to exclude rather than

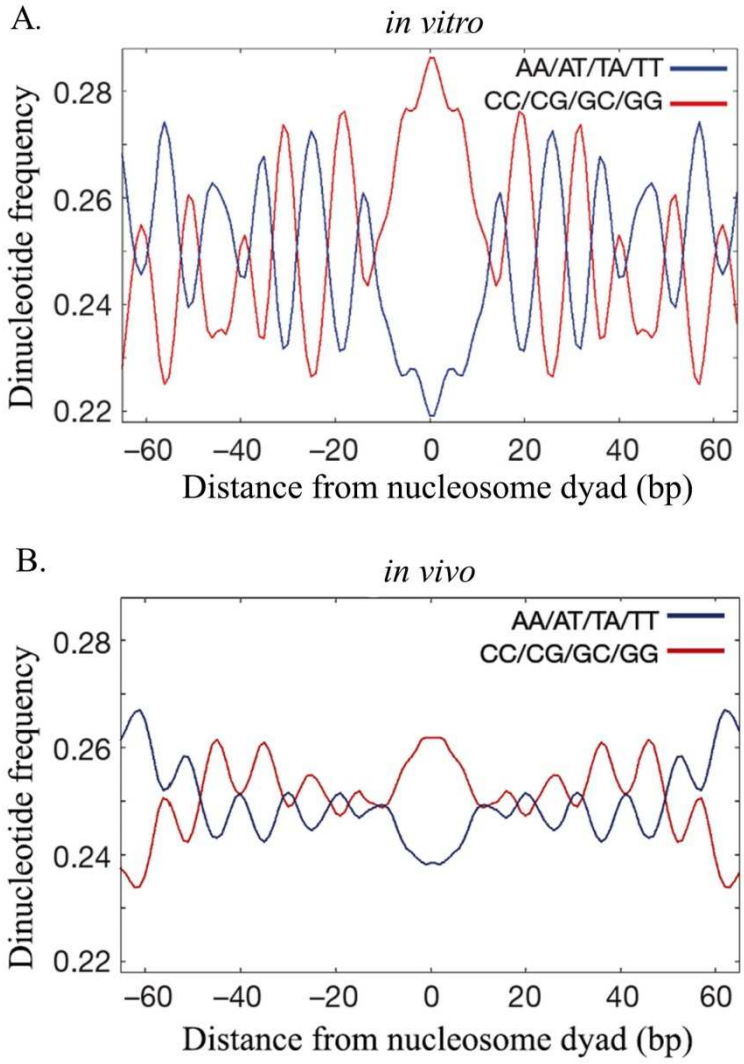


Figure 5: Dinucleotide periodicity of nucleosome sequences in yeast. The fraction of AA/AT/TA/TT and CC/CG/GC/GG dinucleotides at each position of the alignment is shown for A) in vitro map and B) in vivo map (adapted from (Kaplan et al. 2008)).

position nucleosomes. Explicitly, high AT content correlates with low nucleosome occupancy (Hughes & Rando 2009), since the stiffness of poly(dA:dT) sequences strongly disfavor nucleosome formation and hence affects nucleosome depletion *in vivo* (Segal & Widom 2009; Struhl & Segal 2013). Moreover, the presence of the ~10 bp dinucleotide periodicities both *in vivo* and *in vitro* demonstrates that they contribute to the rotational settings of nucleosomes, which define the local orientation of the DNA helix on the histone surface to facilitate DNA wrapping around the octamer. However,

periodic dinucleotides are not sufficient to determine the translational settings of nucleosomes, which refer to the nucleosomal DNA midpoint position relative to genome. Indeed, other *trans*-factors are required for the translational positioning of nucleosomes (Jiang & Pugh 2009; Rando & Chang 2009).

3.2 *Trans*-acting Factors

As discussed above, sequence-driven nucleosome positioning models have suggested that besides DNA sequence, there are other factors contributing to the determination of nucleosome positioning. A clear example about the role of *trans*-factors is provided by the *in vitro* nucleosome reconstitution on the PHO5 promoter. In this study, only DNA and isolated histones were used to reconstitute nucleosomes, which failed to reproduce the *in vivo* nucleosome positioning (Korber et al. 2004). Zhang *et al.* also showed that *in vitro* assembled nucleosomes have limited preferences for translational positions (Zhang et al. 2009). Lastly, different degrees of NFRs *in vivo* and *in vitro* maps have suggested that even though poly(dA:dT) is the major determinant of NFRs, there are other *trans*-acting factors contributing to formation of NFRs (Kaplan et al. 2008).

To date, four major classes of *trans*-factors have been implicated in nucleosome positioning: chromatin remodelers, histone modifications and variants, transcription factors and the RNA polymerase II.

3.2.1 ATP-dependent Chromatin Remodelers

Chromatin remodelers are large macromolecular machines that use the energy of ATP hydrolysis to move or disassemble nucleosomes (Boeger et al. 2008). They affect the location of nucleosomes by altering DNA-histone interactions (Peterson & Logie 2000). Depending on the sequence homology of their ATPase unit, chromatin remodelers are classified into four distinct families: ISWI (ISW1a, ISW1b, and ISW2), INO80/SWR1, CHD, and SWI/SNF (including RSC) (Mellor & Morillon 2004). Among them, ISW1, ISW2, CHD1, and INO80 require linker DNA to reposition nucleosomes, whereas SWI/SNF and RSC can slide nucleosomes without linker DNA (Whitehouse et al. 1999; Whitehouse et al. 2003).

The genome-wide effect of chromatin remodelers on nucleosome positioning was earlier demonstrated for the RSC complex. Two independent studies (Badis et al. 2008; Parnell et al. 2007) have reported that deletion or conditional mutant of RSC lead to increase in the nucleosome occupancy of the RSC-regulated genes, revealing its role in regulation of nucleosomes. Another genome-wide study has shown that in the absence of ISW2 remodeler, +1 nucleosomes shift away from the 5' end NFR (Whitehouse & Tsukiyama 2006), pointing out the effect of ISW2 on +1 nucleosomes. Furthermore, the role of ISW2 on +1 nucleosome positioning is required to inhibit cryptic transcription (Whitehouse & Tsukiyama 2006).

More recently, the role of chromatin remodelers has been directly demonstrated by *in vitro* reconstitution assays (Zhang et al. 2011). Basically, the addition of yeast crude extract and ATP to the reconstituting elements enhances depletion of nucleosomes around promoters to the extent observed *in vivo*, and reproduces *in vivo* -1 and +1 signals. Since this reconstituted map has been able to reproduce the important signals of *in vivo* maps; it has been proposed that chromatin remodelers guide proper nucleosome positioning. The authors have further elucidated the role of chromatin remodelers, suggesting that they may facilitate the nucleosome positioning by packing them against a barrier at 5´ends (Zhang et al. 2011; Yen et al. 2012).

Finally, a detailed study about chromatin remodelers has revealed the nucleosome specificity of each chromatin remodeler, and hence, their specific role in nucleosome positioning (Yen et al. 2012). Basically, using the catalytic or regulatory subunits of chromatin remodelers, they have determined the genome-wide effect of various remodelers on positioned nucleosomes by classifying them based on their nucleosome specific enrichments. Accordingly, ARP5 (subunit of INO80), IOC3 (subunit of ISW1) and ISW2 predominantly interact with +1 nucleosomes, whereas IOC4 (subunit of ISW1) and ISW1 interact with further downstream nucleosomes in the coding region. Finally, INO80, RSC8 (subunit of RSC) and SNF2 (subunit of SWI/SNF) broadly interact with nucleosomes flanking NFRs (Yen et al. 2012). They have further exhibited the directionality of chromatin remodelers in nucleosome dynamics by comparing nucleosome positioning of wild-type and deletion mutant of chromatin remodelers.

Nevertheless, although chromatin remodelers can partially reconstitute the *in vivo* nucleosome pattern and can reposition a portion of nucleosomes, there

are some aspects of nucleosome positioning that remain unexplained, indicating that additional *trans*-factors should play a role.

3.2.2 Posttranslational Histone Modifications and Histone Variants

As mentioned earlier, histone proteins can undergo posttranslational modifications on their tails, which might alter their interaction with DNA. Specifically, it was reported that removal of complete histone tail or histone acetylation generates more unstable and accessible nucleosomes (Anderson & Widom 2001; Polach et al. 2000). Those nucleosomes are prone to be dynamic and become fuzzier, highlighting the direct role of histone modifications on nucleosome positioning. However, the changes on nucleosome positioning usually is the consequence of indirect effect of histone modifications. For instance, certain histone modifications help to recruit chromatin remodelers, which in turn modify the nucleosome location, as observed by the co-localization of SWI/SNF complex with high histone acetylation nucleosome containing regions (Hassan et al. 2001).

Moreover, histone variants, which replace core histones, might also alter bonds and interactions with DNA, affecting nucleosome positioning (Guillemette & Gaudreau 2006; Zhang et al. 2005). In yeast, there are two histone variants; the H2A variant H2A.Z and the H3 variant Cse4 (chromosome segregation 4). Cse4 is located at centromeres and has a function in centromere structure. H2A.Z is usually found around 5' end NFR and is mostly enriched at well-positioned -1 and +1 nucleosomes (Mavrich et al. 2008). It has been reported that well-positioned H2A.Z containing nucleosomes might contribute to the statistical positioning of nucleosomes (Mavrich et al. 2008; Raisner et al. 2005). Furthermore, involvement of H2A.Z in transcription activation has been widely studied (Zlatanova & Thakar 2008) which involves rapid eviction of unstable H2A.Z containing nucleosomes. Finally, *in vitro* studies have exhibited that H2A.Z exchange might cause nucleosome sliding to a more stable position (Guillemette et al. 2005).

3.2.3 Transcription **Factors**

TFs are proteins that bind to specific DNA sequences to regulate gene expression. TFs can directly influence nucleosome positioning by competing with histone cores for binding to DNA. Consistently, global measurements of *in vivo* nucleosome positions and *in vitro* reconstitutions have revealed that most of the transcription binding sites (TFBSs) reside in nucleosome depleted

regions (Yuan et al. 2005; Albert et al. 2007). This competition depends on both relative affinity of TFs to DNA sequence and the abundance of TFs. In particular, the binding sites of relatively more abundant Abf1 and Reb1 TFs are more nucleosome depleted *in vivo* than *in vitro* nucleosome maps (Kaplan et al. 2008). Even if the detailed mechanism of such competition is elusive, one hypothesis involves the recruitment of chromatin remodelers, which in turn evict nucleosomes (Hartley & Madhani 2009; Choi & Kim 2008). Indeed, it has been demonstrated that there is an increase in the nucleosome occupancy of a NFR subset in conditional mutants of Abf1 and Reb1, which was also observed in the conditional mutant of RSC chromatin remodeler (Hartley & Madhani 2009).

3.2.4 RNA Polymerase II

As ~ 2/3 of the yeast genome encodes for proteins and thus experiences the RNA polymerase passage, the effect of RNA polymerase on chromatin structure cannot be negligible. RNA polymerase seems to require disruption of DNA-histone interactions, influencing the structure and positioning of nucleosomes (Bintu et al. 2011). Explicitly, there are evidences that RNA Polymerase II (Pol II) elongation affects nucleosome positioning in coding and promoter regions. For example, genome-wide co-localization of CHD1 and ISW1 chromatin remodelers with Pol II suggests that during its passage along DNA, Pol II recruits the chromatin remodeler to slide nucleosomes. Furthermore, transcriptionally incompetent cell extracts have shown poorly positioned nucleosome arrays, supporting the role of Pol II in nucleosome positioning (Hughes et al. 2012; Struhl & Segal 2013). Moreover, a temperature-sensitive mutation of Pol II subunit has demonstrated that upon polymerase loss, the NFR becomes shorter due to a -1 nucleosome higher occupancy (Weiner et al. 2010). This observation is consistent with the reported role of RNA polymerase in -1 nucleosome eviction (Venters & Pugh 2009). The same study has further exhibited that downstream nucleosomes shift away from the NFR in the absence of polymerase activity. This finding proposes that RNA polymerase might shift the nucleosomes toward NFR until +1 nucleosome is as close as possible to poly(dA), in agreement with the statistical positioning model (Weiner et al. 2010). Altogether, these observations emphasize the role of Pol II on nucleosome positioning.

The preinitiation complex (PIC) has also been suggested to have a role in fine-tuning the +1 nucleosome position (Struhl & Segal 2013), due to its

nucleosome-binding subunits and its strong affinity for the TSS (Jiang & Pugh 2009).

Overall, various genetic, biochemical and informatics analyzes have exhibited that genome-wide nucleosome positioning is determined by the combination of *cis*- and *trans*-acting factors (**Figure 6**).

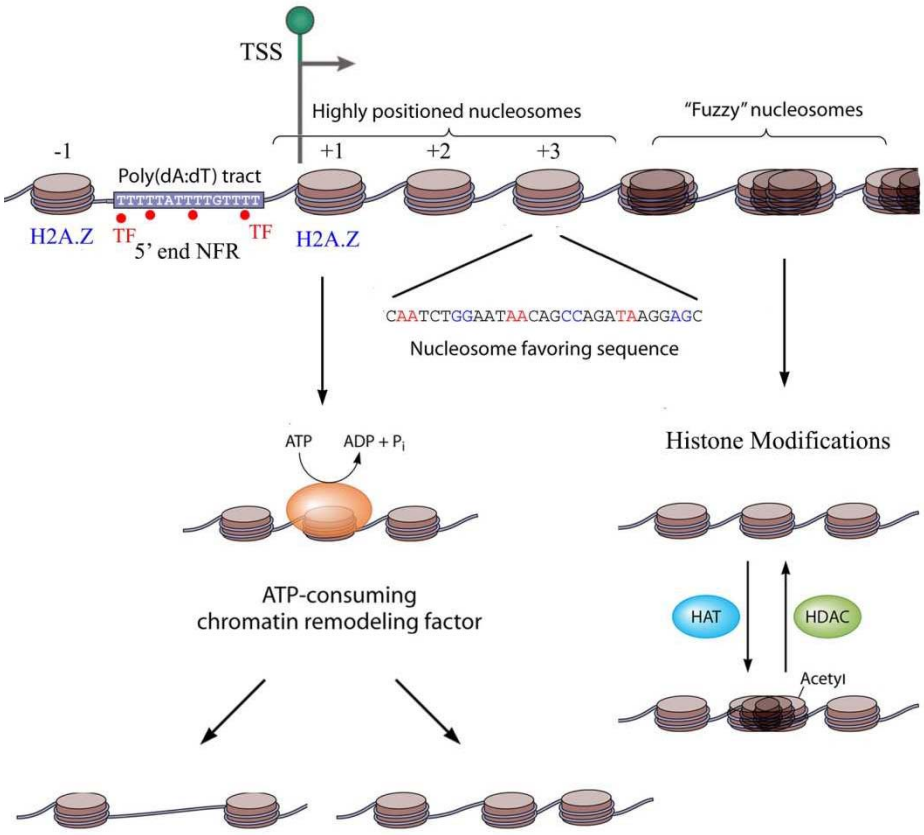


Figure 6: Determinants of nucleosome positioning. A summary of cis and trans factors affecting nucleosome positioning is shown. The most important cis factors are demonstrated as poly (dA:dT) tracts and 10 bp periodicity of dinucleotides which are important for rotational settings. Trans factors like chromatin remodelers can evict or slide nucleosomes, whereas histone modifications alter the accessibility of nucleosomes. Transcription factors indicated as red circles and H2.A.Z. variant usually affect the nucleosome positioning around TSS. (modified from (Jansen & Verstrepen 2011)).

3.3 Variations in Nucleosome Patterns

Nucleosome patterns, influenced by a combination of diverse factors, are representative for gene-averaging. Therefore, the accurate pattern at the individual gene level may vary and also the relative contribution of the acting factors to nucleosome positioning (Struhl & Segal 2013), depending on the promoter type, transcription status and cell cycle process.

Before discussing the divergence in nucleosome patterns along the genome, it is noteworthy to point out the variations on nucleosome maps derived from MNase digestion. Weiner *et al.* have analyzed the effect of MNase digestion level on the resulting nucleosome maps (Weiner et al. 2010). Despite the extensive similarities between maps, they detected particular differences. Accordingly, -1 nucleosome appeared to be the most sensitive to digestion levels, suggesting a weaker binding (Weiner et al. 2010). Moreover, NFRs at 3' and 5' gene ends were wider in higher digestion levels, pointing out that some NFRs might actually correspond to easily digested nucleosomes that are easily removed (Weiner et al. 2010). These findings highlight the importance of considering the applied level of MNase digestion when comparing distinct nucleosome maps.

Accurate analyzes of genome-wide nucleosome maps have also demonstrated that the nucleosome pattern around TSSs might vary depending on the transcription status and promoter architecture (Lee et al. 2007; Jiang & Pugh 2009; Zaugg & Luscombe 2012). Different studies have shown that highly and constitutively expressed genes, i.e. those encoding for ribosomal proteins, tend to have lower -1 nucleosome occupancy and wider 5' end NFR (Lee et al. 2007; Weiner et al. 2010; Radman-Livaja & Rando 2010), leaving the promoter region free of nucleosomes to facilitate transcription machinery binding. Conversely, the promoter region of stress-related genes are more occupied by nucleosomes, leading to a less defined 5' end NFR (Albert et al. 2007; Cairns 2009). Furthermore, Tirosh and Barkai have classified the promoters into two broad classes depending on their nucleosome pattern: open and closed (**Figure 7)** (Tirosh & Barkai 2008). "Open" promoters mostly represent the canonical nucleosome pattern: wider NFR flanked by H2A.Z containing well-positioned -1 and +1 nucleosomes (Cairns 2009; Tirosh & Barkai 2008). The genes in that class are usually constitutively expressed with little variation in expression, i.e. with lower transcription noise. Conversely, "closed promoters" tend to have a

more irregular pattern, with higher nucleosome occupancy around TSSs, which might be related to the dependency on chromatin remodeling complexes (Basehoar et al. 2004). These promoters usually contain TATA boxes and they are characterized by higher transcription plasticity (Albert et al. 2007; Tirosh & Barkai 2008).

More recently, Zaugg and Luscombe have broadened the promoter-dependent classification by an upgrade to a four-state model of nucleosome architecture around TSSs (Zaugg & Luscombe 2012). In addition to NFR status, open or closed state, they include the configurations of -1 and +1 nucleosomes in correlation with gene expression into promoter classification. Accordingly, the presence or absence of TATA box usually determines the open or closed state, as indicated in the previous studies (Cairns 2009; Tirosh & Barkai 2008). Moreover, +1 nucleosome is highly occupied and well-positioned in transcriptionally active genes, whereas it is fuzzier in transcriptionally inactive genes. -1 nucleosome shows a stable configuration for transcriptionally active genes with low expression noise. However, TATA-box containing genes, regardless of their transcription status, generally have fuzzy or poorly positioned -1 nucleosomes (Zaugg & Luscombe 2012). Although this study has been able to show that the nucleosome pattern determines the transcription state and transcription noise, it has been unable to find any association between expression levels and nucleosome occupancy (Zaugg & Luscombe 2012).

These studies altogether indicate that even though most of the genes follow a canonical nucleosome pattern, some of genes do show variations in their nucleosome organizations depending on their promoter type or transcriptional status, pointing out the interplay between nucleosome organization and transcription. Additionally, the nucleosome pattern shows some variations throughout the cell cycle, especially during DNA replication, the details of which will be discussed later in this thesis.

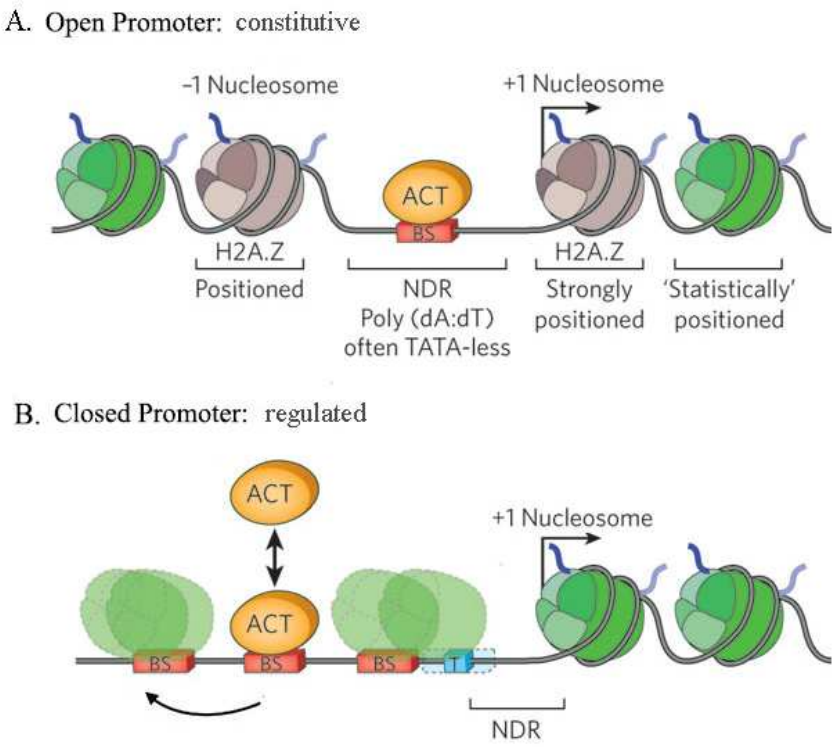


Figure 7: Promoter architecture. A) Open promoters have a nucleosome free region (NFR) adjacent to transcription start site (TSS) flanked by H2A.Z containing well-positioned nucleosomes. This architecture is usually seen at constitutive genes. B) Closed promoters usually have narrower and covered NFR. They are highly regulated by chromatin remodeling complexes. Transcription factor binding sites (TFBS) for transcriptional activators (ACT) are shown as red boxes (taken from (Cairns 2009)).

4 Correlation between Chromatin Structure and Transcription

Chromatin structure has a major impact on gene regulation, since it affects the accessibility of DNA to regulatory factors. It was believed earlier that the presence of nucleosomes directly inhibits transcription by preventing the binding of transcription machinery (Morse 1989). The first *in vivo* evidence for the role of chromatin in transcription regulation was provided by the studies of altered histone expression levels. In particular, depletion of H4 histone, which results in nucleosome loss at PHO5 promoter, changes the transcription state of PHO5 from repressed to active (Han & Grunstein 1988). In consistency with this observation, it was illustrated that highly expressed genes have lower

nucleosome occupancy *in vivo* (Lee et al. 2007; Rando & Chang 2009). These findings and various studies have suggested nucleosomes as negative regulators of transcription (Lee et al. 2007; Rando & Chang 2009; Morse 1989; Han & Grunstein 1988; Kornberg & Lorch 1999). Thus, for the activation of transcription, some DNA-histone interactions should be disrupted or modified for higher accessibility. Histone modifying enzymes, histone variants and chromatin remodelers facilitate and contribute to modify nucleosomes for the transcription activation, which in turn define transcriptional chromatin marks in the genome. For example, histone acetylation has been linked to active transcription, since acetylation relaxes nucleosomal DNA into a more open conformation to enable the movement or remodeling of nucleosomes (Clayton et al. 2006). Indeed, histone acetylases (HATs) and acetylated H3 and H4 are enriched at active promoters (Pokholok et al. 2005). In addition, genome-wide studies have shown that H2A.Z is preferentially enriched at promoters that are poised for transcription activation (Zhang et al. 2005). Once the promoter is activated, H2A.Z is rapidly disassociated to facilitate RNA polymerase passage. After transcription is completed, it is again assembled at +1 nucleosome (Rando & Chang 2009). Thus, due to its rapid turnover, H2A.Z is suggested as a general transcription activator that is incorporated during repression and evicted when promoter activation is required. Lastly, ISW2 chromatin remodeler has been proposed as a transcription repressor, since it regulates transcription by sliding nucleosomes toward promoters, hence preventing the access of RNA polymerase to the promoters.

Other studies have reported conflicting observations, suggesting a more complex role of nucleosome positioning on transcriptional activity. An early study demonstrated that the global depletion of nucleosomes changes the expression of $1/4^{th}$ of yeast genes. However, nearly half of them show a decrease in the expression rather than an increase, in contrast to what had been suggested before (Wyrick et al. 1999). Consistently, distinct nucleosome maps generated from cultures grown under different carbon sources, before and after cell heat shock, or before and after glucose-induced transcription do not exhibit general and observable alterations at promoters upon transcriptional changes, despite some local changes (Zawadzki et al. 2009; Shivaswamy et al. 2008; Kaplan et al. 2008). In summary, the correlation between nucleosome positioning and transcription is probably more complex than what simple models suggest.

5 Yeast Cell Cycle and Coupled Chromatin Dynamics

As mentioned before, there are several factors affecting chromatin structure. Cell cycle is one of the factors that determine the overall chromatin organization and hence, nucleosome positioning.

5.1 Basic Properties of the Cell Cycle

Cell cycle or cell division cycle is the chain of events where cell replicates all of its components and divides into two daughter cells. In eukaryotic cells, the cell cycle consists of four distinct phases: G1, S, G2 and M. G1, S and G2 phases are also referred as interphase, while M phase stands for mitosis (**Figure 8**).

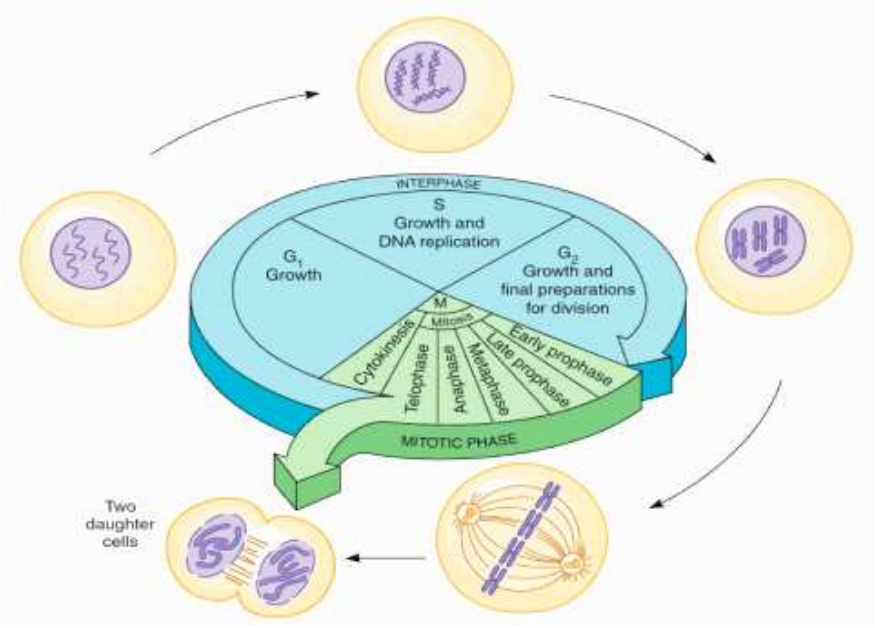


Figure 8: Eukaryotic cell cycle. The diagram shows the progression of a cell through mitosis to produce two daughter cells, modified from http://www.pitt.edu/~super1/lecture/lec19281/004.htm

Gap 1 (G1) phase, is the first stage of interphase. It covers the period from the end of the former M phase until the beginning of the DNA synthesis. During G1, cell grows with high biosynthetic activities in preparation for DNA replication and it reviews whether the conditions are appropriate before the synthesis (S) phase starts. In S, DNA is replicated to create identically duplicated chromosomes. During this phase, the amount of DNA is doubled

and the cells grow further. The proteins necessary for this phase are continuously synthesized. Once the whole chromosomes are replicated, the cell enters to the next gap 2 (G2) stage. Like G1 phase, G2 is another gap phase to ensure that the cell is ready to proceed in the cell cycle. The cell continues growing until it enters into mitosis (M) phase to have the final cellular size. M is the nuclear division phase where the cell divides its duplicated chromosomes into two identical sets in two different nuclei. Mitosis consists in turn of five distinct sequential phases; prophase, metaphase, anaphase, telophase and cytokinesis. After the completion of every cell cycle, each cell contains all the required information and machinery to repeat the process again (Murray & Hunt 1993).

The duration of the cell cycle varies from organism to organism and from cell to cell. It can be as short as 8 minutes, like in some fly embryos, or it can last up to a year, as it occurs in certain liver cells. Most of the variations in cell cycle length are usually due to whole chromosome replication times; in general, the smaller the genome is, the faster it replicates its DNA. In *S. cerevisiae*, the cell cycle time is around 90 minutes (Brewer et al. 1984)

5.2 Cell Cycle Regulation in *S. cerevisiae*

The cell cycle underlies all biological growth and development in living organisms. Therefore, this vital process is under a tight regulation to assure a successful division and efficient proliferation. Nurse, Hartwell and Hunt discovered two regulatory proteins, cyclin and cyclin dependent kinase (CDK), which control the cell cycle progress (Nurse et al. 1976; Hartwell et al. 1970; Evans et al. 1983). These proteins are the so-called checkpoints, since they review and verify whether each cell cycle phase is accurately completed and whether the cells can enter to the next stage. Their expression levels or activation states usually fluctuate throughout cell cycle. More recently, around 800 genes have been detected with differential expression throughout the yeast cell cycle, based on a large-scale array transcriptome analysis (Spellman et al. 1998). However, the number of the key regulators that are controlling the cell cycle is much smaller. Overall, the yeast cell cycle regulatory network can be classified into four groups; cyclins (Cln1-3, Clb1-6); CDK inhibitors (CDIs) (Sic1, Cdh1, Cdc20, Cdc14); heterodimeric transcription factors SBF (Swi4 and Swi6) and MBF (Mbp1 and Swi6); and checkpoints (Li et al. 2004).

Cyclin proteins control the progression of the cell cycle by activating CDK enzymes. As the name implies, the abundance of the cyclins varies in a cyclical fashion throughout the cell cycle, hence inducing oscillations in CDK activity (Morgan 2007; Evans et al. 1983). In budding yeast, around 20 cyclins bind to one of the five different CDKs, Cdc28, Pho85, Kin28, Srb10 or Ctk1 (Andrews & Measday 1998). However, only Cdc28 (also termed CDK1), which is activated by 9 cyclins, has a clear role in the cell cycle regulation, while the others are involved in transcription regulation. Cdc28 forms an active heterodimer with other cyclins, being CDK1 the catalytic subunit. The active CDK1 phosphorylates the target proteins, which in turn stimulate the entry into the subsequent phase of the cell cycle. Different cyclin-CDK1 heterodimers have different targets; thus they control different stages of the cell cycle.

CDIs physically inhibit the activity of CDK1 and thus act as negative regulators of cell cycle progression. Once the cell enters to G1 phase, the inhibitors prevent CDK1 functioning from acting too long or too short. As the cell proceeds to S phase, these inhibitors are degraded.

SBF and MBF transcription factor classes activate certain genes during the G1/S transition. MBF-regulated genes generally control DNA replication, whereas SBF-regulated genes are involved in cell morphogenesis and spindle pole body duplication.

Checkpoint proteins are used to monitor and verify the progress over the cell cycle. When damaged or incomplete DNA is detected, the checkpoints are activated and the cell cycle is arrested until the damage is repaired. G1 (restriction), G2 and metaphase checkpoints are the major inspection sites to assure the fidelity of the cell cycle.

5.3 Cell Cycle Phases

5.3.1 G1 Phase (Start)

Budding yeast cells enter the cell cycle depending on their size and the environmental conditions such as the availability of nutrients and the presence of a mating pheromone. Once they are suitable, the cells commit irreversibly to initiate the cell cycle at late G1 phase, referred to as the START point. Events at START depend on the activation of Cdc28 that is controlled by Cln3 cyclin at G1 phase. At the beginning of G1 phase, Cln3 binds to Cdc28 and forms an active complex (**Figure 9**). The activated Cdc28 phosphorylates and inactivates Whi5, which is the inhibitor of SBF and MBF transcription factor complexes (Breeden 1996). This event triggers the liberation of SBF and MBF complexes from Whi5 (de Bruin et al. 2004). SBF and MBF complexes in turn activate G1-specific genes like Cln1, Cln2, Clb5 and Clb6 cyclins that form complex with Cdc28 to enable cell cycle progression. While Cln1, Cln2 /Cdc28 complexes are active; Clb5, Clb6/Cdc28 complex activity is detained by Sic1 inhibitor. Cln1, 2/Cdc28 complexes drive the progression through START by initiating the processes leading to bud formation and DNA replication. Furthermore, they phosphorylate and block the Sic1 inhibitor to liberate Clb5, Clb6/Cdc28 complexes. Clb5 and Clb6 are the first activated B type cyclins, which control the progression of cell cycle into and out of mitosis. Their activity rises at late G1 and their expressions stay high over S and M phases. Clb5-6/Cdc28 active complexes directly promote initiation of DNA replication. Moreover, they further trigger the phosphorylation, and hence the degradation, of Sic1 inhibitor. Thus, Clb5 and Clb6 control their own activation in a positive feedback loop. Finally, the degradation of Clb5 and Clb6 is usually mediated by the APC/Cdc20 complex.

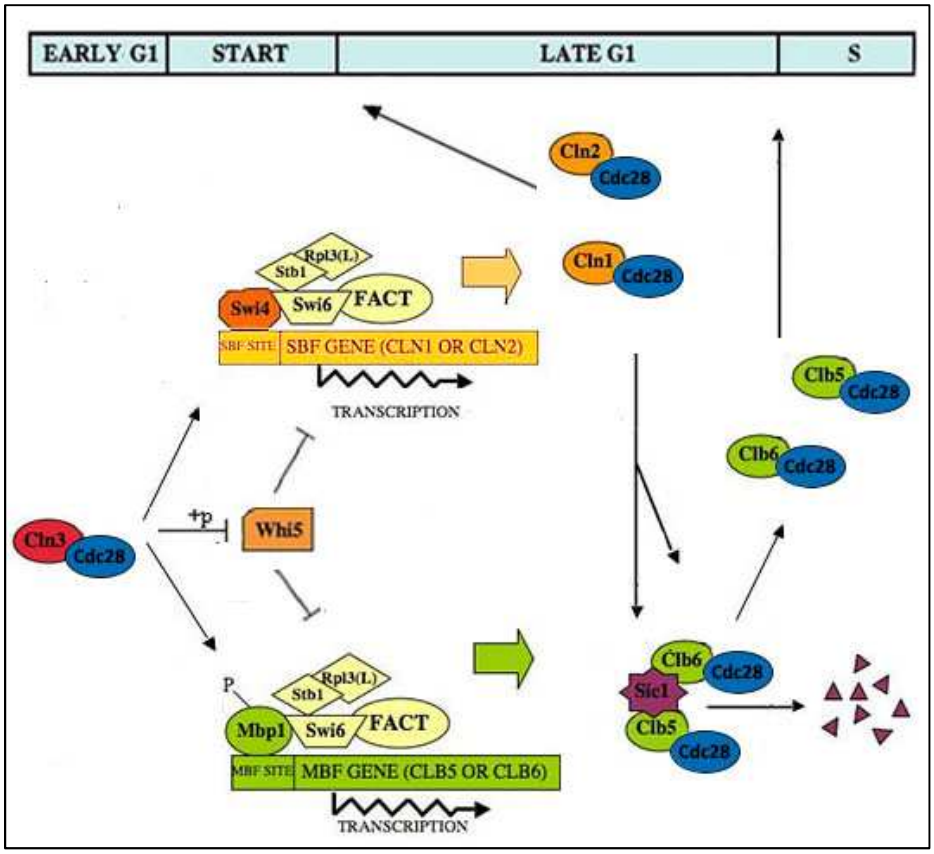


Figure 9: G1/S transition in S. cerevisiae. At G1 phase, Cln3 triggers deactivation of Whi5, which in turn liberates SBF and MBF-dependent genes. These genes are G1-specific, like Cln1, Cln2. Clb5 and Clb6, which are required for G1/S transition. (Modified from (Douglas Maya 2011)).

5.3.2 S Phase (DNA Replication)

In eukaryotic cells, DNA replication starts from multiple specific regions of the genome, called replication origins. Replication origins were first characterized in *S. cerevisiae* and also referred to as autonomously replicating sequences (ARSes). In yeast, there are 460 estimated ARSes from which bidirectional replication is initiated (Wyrick et al. 2001). They are characterized by their AT-rich content, where the unzipping of DNA is easier due to the relatively fewer hydrogen bonds between base pairs, which is critical for origin function (Gerard & Gluzman 1986).

Replication is initiated with the assembly of the pre-replicative complex (pre-RC) in a highly ordered stepwise process that starts with recognition of ARSes

by the origin recognition complex (ORC). ORC specifically recognizes and binds to 11 bp DNA consensus sequence within the origins, called ARS consensus sequence (ACS). ORC is constitutively bound to ACSes throughout the cell cycle (Aparicio et al. 1997; Bell & Stillman 1992). The next step of pre-RC is the recruitment of Cdc6 and Cdt1 by ORC (**Figure 10**). Cdc6 is activated at G1 phase and inactivated at S by the Clb5-6/Cdc28 complex, so that pre-RC cannot be assembled outside of G1 phase. Thus, by controlling Cdc6 activity, the Clb5/Cdc28 complex assures that DNA replication only occurs once per cell cycle. The binding of Cdc6 and Cdt1 to ORC is essential for the recruitment of the next component of pre-RC, the six-subunit mini-chromosome maintenance (MCM) protein complex. MCM complex functions as helicase to unwind DNA prior to DNA replication and it also has a role in both the initiation and elongation steps of DNA synthesis.

At the G1/S transition, the pre-RC is then transformed into an active replication fork. For the assembly of the active replication fork, Cdc-6 is disassembled from the pre-RC under the regulation of Clb5-6/Cdc28, while MCM complex is phosphorylated by the Dbf4/Cdc7 heterodimer. Upon phosphorylation, MCM interacts with a component of the replication fork, Cdc45 (Gregan et al. 2003), which is crucial for the recruitment of other replication proteins such as GINS complex, DNA polymerase α (pol α), DNA polymerase ε, replication protein A (RPA) and proliferating cell nuclear antigen (PCNA) onto the chromatin (Aparicio et al. 1999). The binding of these replication proteins establishes the initiation complex, which coordinates DNA replication. It is initiated by Pol α, which synthesizes short RNA primers for leading and lagging strand synthesis. However, due to the low processivity and the lack of proofreading ability, Pol α disassociates after initiating DNA synthesis and Pol ε and Pol δ take over. Due to the antiparallel structure of the duplex DNA, DNA synthesis occurs in opposite directions between the two new strands at the replication fork. Pol ε continuously synthesizes DNA on the leading strand in the direction of DNA unwinding, whereas pol δ synthesizes DNA on the lagging strand in a fragmented or discontinuous manner. These discontinuous 100-200 bp-long fragments are the so-called Okazaki fragments (Sakabe & Okazaki 1966).

As DNA replication is initiated on multiple origins, replication forks meet and terminate DNA synthesis at multiple points. However, DNA replication cannot

reach the very end of the chromosomes but stops at repetitive telomeric regions, shortening the telomeres of the daughter strands.

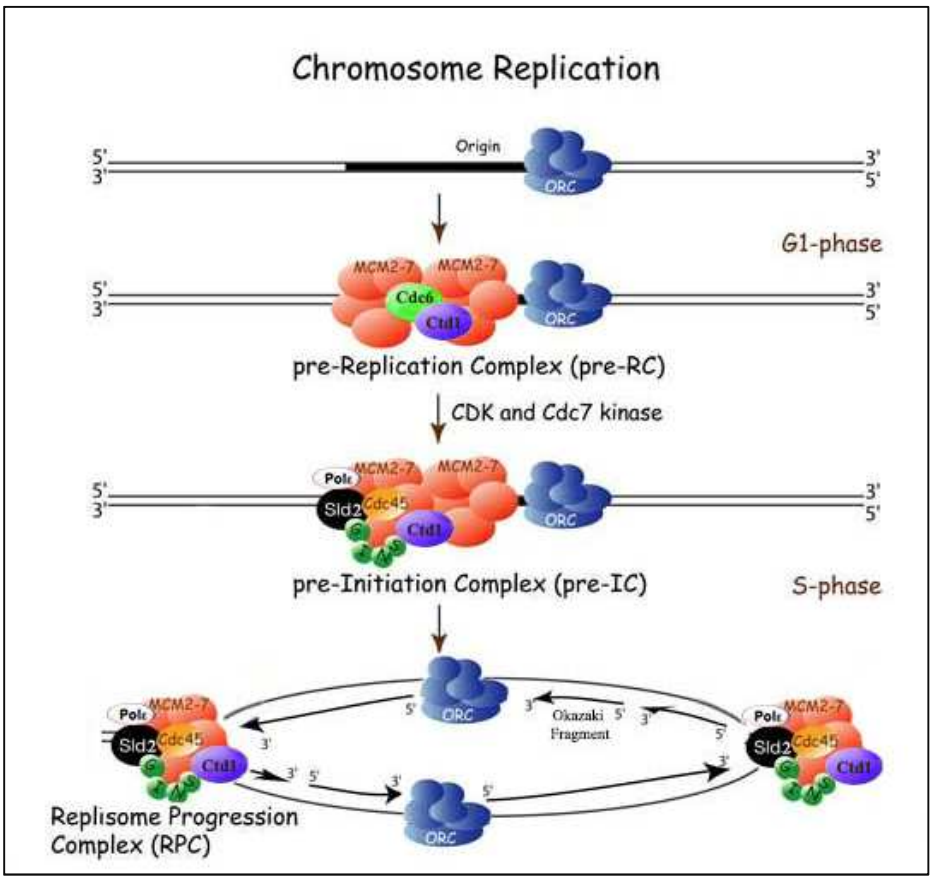


Figure 10: The Replication complex assembly in *S. cerevisiae*. Simplified diagram shows the recruitment of replication proteins onto replication origin and origin firing. Modified from http://www.paterson.man.ac.uk/images/cell%20cycle/fg_replication_large_new.jpg

5.3.3 G2 and M Phases

Once DNA replication is completed, entry to G2 and mitosis is driven by other B type cyclins, Clb1-4, and ubiquitin-dependent proteolysis, which is governed by the anaphase-promoting complex/cyclosome (APC/C). Clb1-4 accumulation is enabled by the high levels of Clb5 and Clb6 during S phase to form active Clb1-4/Cdc28 complexes, which have a role in the intra-nuclear spindle assembly (Fitch et al. 1992). These complexes also stimulate the activation of APC/C complex at anaphase, which in turn mediates its own proteolysis as well as other M phase regulatory proteins like securins, polo kinases and spindle-

associated proteins (Irniger 2002; Rudner & Murray 2000). The activity of APC/C is controlled by the association of Cdc20 and Cdh1 proteins (**Figure 11**). At metaphase, when sister chromatids are aligned, APC/C/Cdc20 complex is activated, which liberates Esp1 (separase) by degrading its inhibitor Pds1 (securin), leading in turn to the sister chromatid separation (Zachariae & Nasmyth 1999; Nasmyth et al. 2000). Once the sister chromatids are separated, they are pulled to opposite poles of the cell at anaphase. Another APC/C activator, Cdh1, is dephosphorylated by Cdc14 phosphatase, which is released from nucleolus at telophase upon the stimulation of Cdc5 and Cdc20/APC/C complex (Pereira et al. 2002; Stegmeier et al. 2002). Consequently, Cdh1/APC/C complex is activated and directly targets the cyclins for degradation. Moreover, Cdc14 dephosphorylates and consecutively activates Cdk inhibitor Sic1 (Zachariae 1999; Visintin et al. 1998), leading to complete elimination of Cdk activity at telophase. As the Cdk/Cyclin complexes are destroyed, the cells properly complete their division and are able to exit mitosis to restart the next cell cycle.

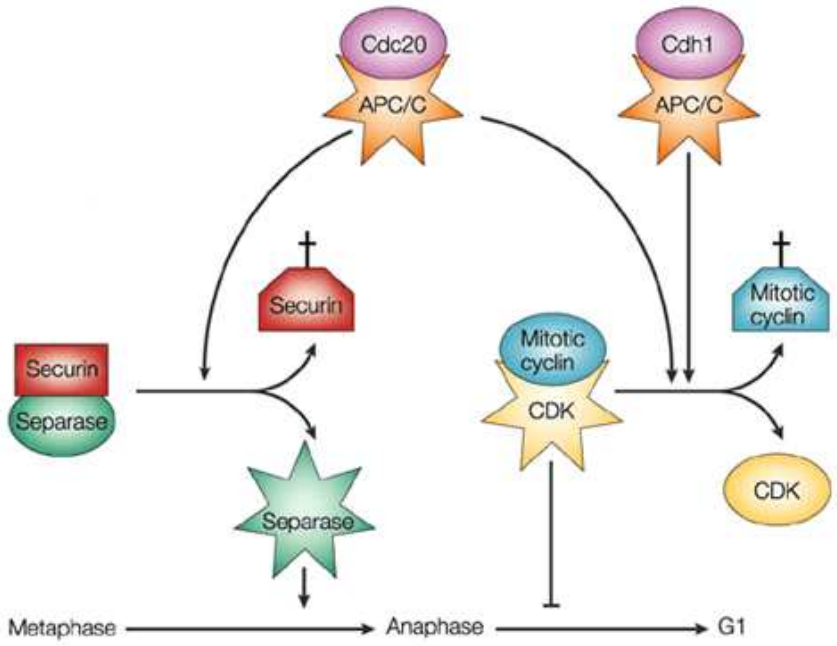


Figure 11: Mitosis control by anaphase-promoting complex/cyclosome (APC/C). The activity of APC/C is controlled by the association of Cdc20 and Cdh1 proteins. APC/C/Cdc20 complex liberates Esp1 (separase) by degrading its inhibitor Pds1 (securin), leading in turn to the sister chromatid separation, whereas Cdh1/APC/C complex directly targets the cyclins for degradation (adapted from (Bardin & Amon 2001)).

5.4 Chromatin Assembly Coupled with DNA Replication

The progression of the cell cycle requires partial unraveling of local chromatin structure to regulate DNA metabolism. This partial unraveling is achieved by the disruption of histone-DNA contacts and can be mediated at two main levels: via ATP-dependent chromatin remodeling factors or via post-transcriptional modifications of the histone tails, as mentioned before (Boeger et al. 2003). Especially during replication, some specific histone-DNA interactions need to be disrupted and reassembled for a faithful DNA replication and proper chromatin structure.

As mentioned before, DNA replication starts with a step-wise formation of pre-replicative complex, which is pioneered by the assembly of ORC at replication origins. It has been postulated that the activation of the origins are associated with some discrete genomic features, i.e. in human genome origins are clustered in GC-rich regions and overlap transcriptional regulatory elements (Cadoret et al. 2008), pointing out the possible role of surrounded chromatin structure and modifications. The details of the link between chromatin state and replication origins will be discussed in greater detail in section 6. Once the origins are activated, the replication fork moves bi-directionally from each origin. The passage of replication disrupts DNA-histone contacts and hence, previously existing nucleosomes as well (Gruss et al. 1993). However, the nucleosomes are not completely disassembled from replicating DNA; rather the parental histones remain associated with one of the strands in a random way.

During DNA replication, parental histones are segregated randomly to both strands (Krude 1999), which transmit the epigenetic information to the newly assembled chromatin. Immediately after DNA replication, new nucleosomes should be assembled onto the nascent chromatin for the complete duplication of the chromatin. The half-nucleosomes with parental histones on the nascent chromatin are completed by newly synthesized histones to form new nucleosomes. Firstly, H3 and H4 histones associate with each other to form H3/H4 tetramer, driven by the anti-silencing factor 1 (ASF-1) (Tyler et al. 1999) (**Figure 12**). Right after their synthesis, ASF-1 stimulates the acetylation of H3/H4 tetramer by Rtt109 or Gcn5 acetyltransferases (Fillingham et al. 2008). Once the H3/H4 tetramer is acetylated, ASF-1 specifically delivers them to the chromatin assembly factor 1 (CAF-1), which consequently recruits the

acetylated histones onto the newly synthesized nascent DNA strand (Smith & Stillman 1989). Yet, another histone chaperone, Rtt106, is implicated in depositing H3/H4 tetramers onto DNA (Huang et al. 2005). The chromatin assembly takes place right after DNA replication and these processes are coupled. This coupling is driven by the physical interaction of CAF-1 with the PCNA component of the replication machinery. PCNA marks newly replicated DNA and recruits CAF-1, which affects epigenetic inheritance of DNA and chromatin structure (Krude 1995; Zhang et al. 2000). Briefly, the proper chromatin assembly coupled to DNA replication requires ASF-1, CAF-1 and Rtt106 histone chaperons and the sliding clamp, PCNA (**Figure 12**).

After the incorporation on the nascent chromatin, the newly synthesized H3 and H4 histone proteins are deacetylated to preserve the parental situation.

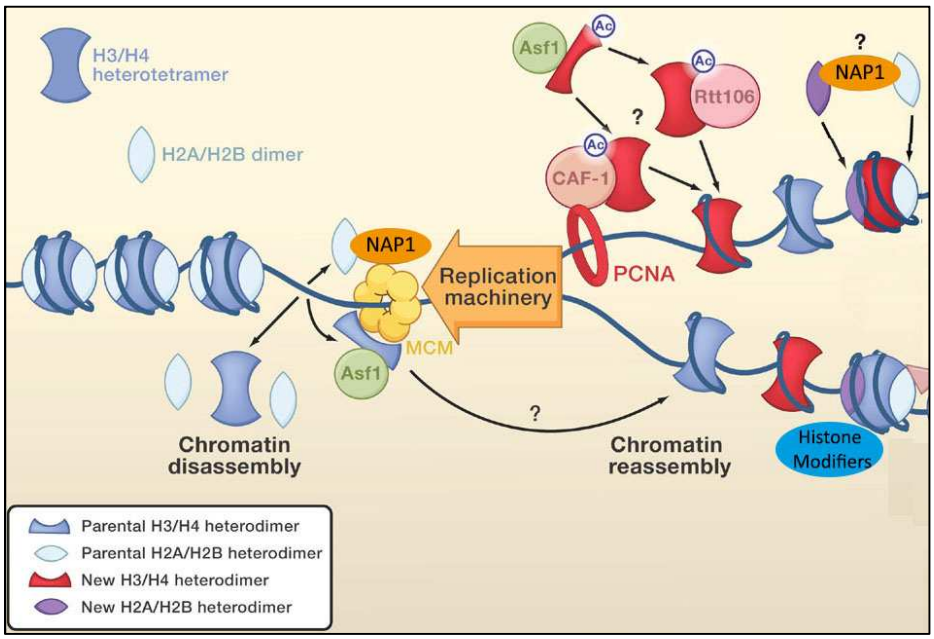


Figure 12: Chromatin assembly during DNA replication. Nucleosome disassembly before DNA replication and new nucleosome assembly driven by histone chaperons are shown. Question marks correspond to the speculative steps (adapted from (Ransom et al. 2010)).

In most of the higher eukaryotes, deacetylation is also crucial to repress the transcription of many genes, so that right after the DNA synthesis, replicated genomic regions are not directly transcribed. However, in *S. cerevisiae*, where a large fraction of the genome is transcribed and a considerable part of the histones are acetylated (Davie et al. 1981), it is not clear whether

deacetylation after DNA replication is crucial for *S. cerevisiae*. Yet, in budding yeast, there are genes coding histone deacetylase subunits, like RPD3, HDA1, HOS1, HOS2 and HOS3. Nevertheless, whether they have a direct role in deacetylation of the newly synthesized histones remains unclear (Verreault 2000).

Once H3 and H4 histones are assembled to the newly synthesized DNA, histone H2A/H2B dimers are then incorporated to complete nucleosome formation (Park et al. 2005). Several studies in higher eukaryotic cells suggest the potential role of the nucleosome assembly protein (NAP1) in deposition of H2A/H2B histones. *In S. cerevisiae*, the strongest evidence comes from biochemical and genetic studies, suggesting its importance in cell cycle progression. NAP1 is required for the regulation of protein kinases like Cdc28-Clb2 and Gin2 during mitosis (Altman & Kellogg 1997). However, direct evidences about its role in nucleosome assembly are not yet reported (Verreault 2000).

6 Chromatin Organization at Replication Origins

Many independent and complementary studies *in vivo* have identified various active replication origins in *S cerevisiae*, which are reported in the OriDB database (Nieduszynski et al. 2007). Even though the 460 identified replication origins follow a common protein loading and activation mechanism, each of them has distinct origin firing times and efficiencies. Although most of the origins are efficient, some of them show lower firing efficiency within a population (Weinreich et al. 2004). Further, while some origins fire right after the cells enter to S phase, referred to as early origins; certain origins fire towards the end of S phase, referred to as late origins. The reason why the cells have evolved to have different origin firing times and efficiencies is still elusive, although local chromatin environment seems to be closely linked. For instance, early and late origins take place at distinct genomic locations. Explicitly, early origins are positioned toward the central part of chromosomes and are correlated with open chromatin structure, whereas late origins are usually located at telomeres, suggesting a position-dependent control of replication origins (Raghuraman et al. 2001). Indeed, an early study demonstrated that replication origins in the proximity of silent loci are inactive, whereas they become active when inserted at more favorable

positions (Vujcic et al. 1999). Moreover, two distinct studies from the same group further supported the positioning effect by showing the alterations in origin timing upon modifying the locations of early or late firing origins (Ferguson et al. 1991; Friedman et al. 1996). Recently, 3D structures of the budding yeast genome highlighted the clustering of replication domains in the chromatin (Tjong et al. 2012). Interestingly, based on 3D structure, early firing origins are in close spatial proximity, usually near centromeres. Furthermore, it was demonstrated that forkhead transcription factors (FKH) 1 and 2 are responsible for establishing certain early replication domains by bringing them in close proximity at G1 to regulate their activity (Knott et al. 2012). Finally, a recent study showed that while some replication origins alter their firing time upon the insertion at different position, certain early origins were not affected. Interestingly, these location-independent origins contain specific sequences for FKH1/2 binding, confirming the role of FKH1/2 in the regulation of certain early origins (Lõoke et al. 2013), indicating that even though most of the replication origin activity is chromatin location dependent, there are some origins that are mediated by additional factors.

In addition to chromatin positioning, nucleosomes have been shown to play a crucial role in origin activation as well. It was earlier demonstrated that in ARS1 origin, the core elements are actually free of nucleosomes (Thoma et al. 1984). Consistently, in higher eukaryotes, ORC binding sites are also suggested to be depleted of nucleosomes (MacAlpine et al. 2010). Various studies further confirmed the depletion of nucleosomes on replication origins (Yin et al. 2009; Eaton et al. 2010). Moreover, *in vitro* chromatin assembly assays and *in vivo* assays proposed ORC as a determinant of nucleosome positioning around origins, since the loss of ORC results in a shift in nucleosome positioning around origins (Lipford & Bell 2001). In the presence of ORCs, ORC-binding sites are occupied and hence, nucleosome formation is interfered at replication origins, resulting in nucleosome depletion. The role of the nucleosome depletion on origin activation was reported by Simpson *et al.* (Simpson 1990). Accordingly, the occupation of ARS by a single nucleosome, driven by specific mutation in DNA sequence, interferes with origin activity (Simpson 1990). Finally, detailed nucleosome positioning around origins has been determined relative to ORC binding site on ACS (Eaton et al. 2010). The NFR is positioned asymmetrically at ACS and flanked by well-positioned nucleosomes (Eaton et al. 2010). The NFR is wider than ORC *DNase I* footprint, proposing that ORC is not the only factor responsible for nucleosome

depletion at origins. Later, Berbenetz *et al.* illustrated that the chromatin organization around replication origins is not uniform but highly diverse (Berbenetz et al. 2010). For instance, the origins in the proximity of TSSs have an extra NFR located upstream of the original NFRs. Moreover, late replication origins tend to have unusual NFRs, either very narrow or very wide. Lastly, two studies showed that certain DNA sequences and properties might play a role in determining the NFR and flanking nucleosomes (Berbenetz et al. 2010; Eaton et al. 2010). Overall, these observations indicate that proper and diverse nucleosome positioning is important for origin activation and timing and the positioning might partially depend on *cis*-binding factors.

The diversity of nucleosome organization and the importance of nucleosome repositioning in origin activation brought the possible role of ATP-dependent chromatin remodeling into focus. Indeed, it was shown that in budding yeast, the absence of SWI/SNF remodeler complex interferes with functioning of certain replication origins (Flanagan & Peterson 1999). Moreover, Zhou *et al.* demonstrated that the absence of SWI/SNF complex subunits interferes with MCM3 loading in plasmid replication origins (OriP) (Zhou et al. 2005). These results together pointed out the role of SWI/SNF complex in replication origin activation, probably via repositioning of nucleosomes to facilitate replication factor binding. Recent studies have enlightened the possible role of INO80 and ISW2 chromatin remodelers in replication origins (Vincent et al. 2008; Shimada et al. 2008). The enrichment of INO80 complex at replication origins, prominently at stalled replication forks under replication stress, suggested a stabilizer role of INO80 for stalled replisomes (Papamichos-Chronakis & Peterson 2008). Additionally, a detailed study proposed that INO80 and ISW2 regulate the efficient progression of replication forks and are particularly required for the activation of late replication regions (Vincent et al. 2008). Possibly, these chromatin remodelers can modify nucleosomes ahead of the replication fork to ease its passage or behind the fork, helping to adopt a proper chromatin state.

Lastly, by altering the DNA accessibility, chromatin modifications have been shown to play a role in replication origin activation. Particularly, there is a strong correlation with acetylation state and origin activation in a way that higher acetylation status stimulates more open chromatin structure enabling a favorable environment for recruitment of replication factors. Consistently, histone deacetylases Sir2 and Rpd3 have been reported to negatively regulate

origin activation (Pappas et al. 2004; Vogelauer et al. 2002). Moreover, deletion of Rpd3 result in earlier firing of some replication origins, confirming the role of higher acetylation levels in origin activation (Knott et al. 2009). Furthermore, H3 and H4 acetylation levels are increased at S phase and G2/M phase around origins (Unnikrishnan et al. 2010).

7 The Centromere

The centromere is a specialized region on every chromosome that links sister chromatids via the kinetochore formed during mitosis. Microtubules attach to the kinetochores via bipolar spindles (Cheeseman & Desai 2008) (**Figure 13**). The interaction between microtubules and centromere, mediated by kinetochore, is required for the faithful separation of replicated chromosomes, so that the cell division can proceed to completion accurately (Cleveland et al. 2003).

In eukaryotes, there are two main types of centromeres: point centromeres and regional centromeres. While point centromeres bind to particular proteins, which recognize specific DNA sequences, regional centromeres are packed into heterochromatin and consist of large arrays of repetitive satellite DNA (Pluta et al. 1995). *S. cerevisiae* centromere is the best characterized point centromere and it has a unique, genetically defined structure (Clarke & Carbon 1980) that is more compact and relatively shorter than canonical nucleosomes, about 125 bp in length. On the other hand, regional centromeres are found in most organisms, from fission yeast to humans. In fission yeast, *S. pombe,* centromeres are 40-100 kb long and consist of a non-repetitive central core flanked by two identical inverted repeats, while in metazoans, centromeres are composed of large arrays of tandem repeats, which lack sequence specificity (Verdaasdonk & Bloom 2011).

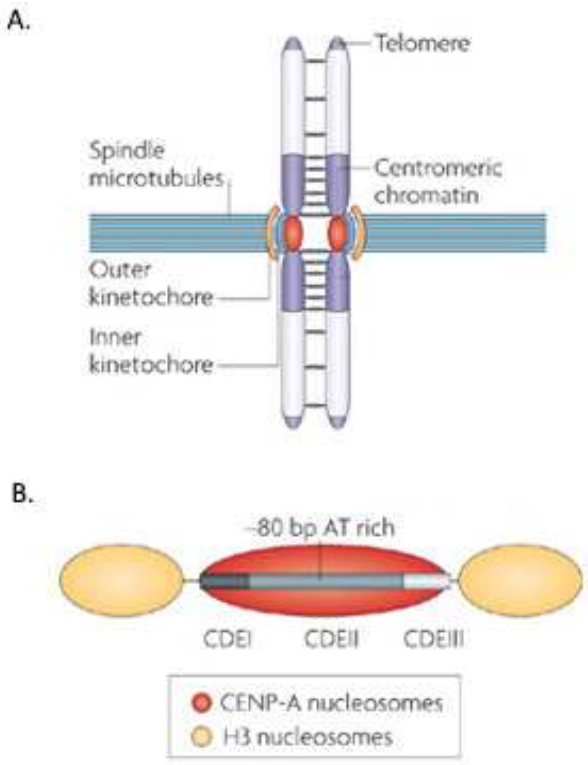


Figure 13: Centromere Structure and Organization. A) Centromeric chromatin links sister chromatid via kinetochore. B) Schematic representation of centromeric DNA in *S. cerevisiae*. Adapted from (Allshire & Karpen 2008)

7.1 Centromeric DNA and Centromere-associated Proteins in Budding Yeast

Centromeric DNA of budding yeast is composed of a non-conserved AT-rich core segment, centromere DNA element II (CDEII, 78-86 bp), flanked by two consensus sequences, CDEI (8 bp) and CDEIII (25 bp) (Ishii 2009; Fitzgerald-Hayes et al. 1982) (**Figure 13**). CDEI is required for high fidelity chromosome segregation, whereas CDEIII is essential for kinetochore formation and centromere activity (McGREW et al. 1986). CDEII is the region where the histone variant, centromeric H3 (CENH3) binds (Bloom & Carbon 1982). The complete deletion of CDEI does not interfere with kinetochore formation but degrades the fidelity of chromosome segregation (Carbon & Clarke 1984). On the other hand, the deletion of CDEII and CDEIII results in the complete loss of centromere activity, and moreover some certain point mutations in CDEIII might affect centromere activity (Carbon & Clarke 1984; Purvis & Singleton

2008). These results demonstrated that CDEII and CDEIII regions are necessary for normal centromere function and CDEI plays a minor role helping to regulate the fine structure.

Studies on centromere proteins have been important to understand centromere function and have allowed the characterization of various proteins localized to CDE regions. Centromere binding factor (CBF) 1 is bound to CDEI region with high sequence specificity. It is postulated that its conserved helix-loop-helix motif might be responsible for this specific binding activity (Cai & Davis 1990). The genetic analysis of CBF1 demonstrated that like CDEI; CBF1 is not essential for centromere function, but it contributes to high fidelity centromere and kinetochore formation (Cai & Davis 1990). Similar to CBF1, CBF3 complex also binds with high specificity to CDEIII element. CBF3 complex consists of four proteins: CBF3a (CBF2, Ndc10, Ctf14), CBF3b (Cep3), CBF3c (Ctf13) and CBF3d (Skp1) (Stemmann & Lechner 1996; Pietrasanta et al. 1999; Lechner & Carbon 1991). All four proteins are essential for cell viability and centromere function. Explicitly, CBF3a, CBF3b and CBF3d are in direct contact with the major groove of CDEIII, but only CBF3b contains a known zinc-finger type DNA-binding motif (Lechner 1994; Espelin et al. 1997). AFM studies show that binding of CBF3 complex results in centromeric (CEN) DNA bending that might facilitate the interaction with other centromeric proteins or correct positioning of the centromeric nucleosome (Pietrasanta et al. 1999; Purvis & Singleton 2008).

The components of kinetochore are other sets of centromere-associated proteins. In *S. cerevisiae*, the kinetochore is composed of around 70 constitutive components and most are conserved (Lampert & Westermann 2011). These components are classified as inner, central and outer domains of kinetochore depending on their location. CENH3, mitotic fidelity of chromosome transmission 2 (Mif2), CBF1 and CBF3 complex together encompass the inner kinetochore, whereas MTW1, CTF19/COMA complexes are the components of central kinetochore (Gascoigne & Cheeseman 2012; Malvezzi et al. 2013). The outer components (NDC80 and DAM1 complex) are located where microtubules are attached. These proteins together with some central kinetochore proteins (Spc105, Mtw1) form a platform for microtubule attachment, which is called KMN network (Bock et al. 2012).

7.2 Centromeric Histone Variant CenH3 and its Deposition at Centromere

It has been reported by many studies that centromeric chromatin has a different characteristic compared to bulk chromatin (Dunleavy et al. 2005; Verdaasdonk & Bloom 2011; Henikoff & Dalal 2005). The earlier findings showed that CenH3 can be co-purified with nucleosome core particles and interact with other histone proteins (Palmer et al. 1987). Moreover, it was demonstrated that CENH3 has histone-like properties of solubility and high identity similarity with H3 (Sullivan et al. 1994). Together, these results suggested CENH3 as a centromere-specific H3 variant in centromeric nucleosome, which is replaced by H3 in the bulk chromatin. Its presence in all the active centromeres characterizes CenH3 as the epigenetic mark that specifies the centromere identity (Warburton et al. 1997). In most organisms, centromeres contain arrays of CENH3 nucleosomes, whereas point centromeres are organized into a single centromeric nucleosomes (Furuyama & Biggins 2007).

The Centromeric histone variant CENH3 is highly conserved throughout eukaryotes and known as CENP-A in humans, centromere identifier (CID) in *D. melanogaster* , CNP1 in fission yeast and Cse4 in budding yeast (Meluh et al. 1998; Takahashi et al. 2000; Henikoff et al. 2000). Cse4 is a 27 kDa protein composed of an N-terminal domain with a unique sequence and a C-terminal domain with a 65% similarity to H3 over its histone fold domain (Glowczewski et al. 2000). The proper deposition of Cse4 at centromeres is enabled by CENP-A targeting domain (CATD) that is composed of loop 1 and the α2 helix within the histone fold, absent in H3 histone (Black et al. 2004). Besides CATD, there are some proteins guiding Cse4 assembly at centromeres. For instance Scm3, a non-histone protein, is required for the correct Cse4 deposition and stability (Stoler et al. 2007). Even though the correct positioning is under tight control, ectopic Cse4 incorporated into euchromatin has also been reported. Explicitly, low levels of Cse4 had been detected in chromosome arms and highly transcribed regions (Camahort et al. 2009; Lefrançois et al. 2009). The mistargeted Cse4 is quite unstable and not sufficient for kinetochore formation. Furthermore, the over-expression of Cse4 does not lead to mislocalization, suggesting the possibility that additional mechanisms exist to remove the exclusively located Cse4 (Crotti & Basrai 2004). One of the mechanisms encompasses ubiquitin-mediated proteolysis that controls the

levels of Cse4 by degrading euchromatic Cse4 (Collins et al. 2005). While ectopic Cse4 is degraded, centromeric Cse4 is protected from the proteolysis, ensuring the incorporation of Cse4 only onto centromeres. Additionally, the SWI/SNF chromatin remodeler complex has been reported to also function in the distribution of Cse4 by removing misincorporated Cse4 from the ectopic sites (Gkikopoulos et al. 2011). Lastly, both replication-dependent H3-H4 chaperon CAF-1 and replication-independent H3/H4 chaperon HIR complex have been implicated in prevention of ectopic Cse4 by evicting those (da Rosa et al. 2011).

7.3 Architecture of Centromeric Chromatin

The structure and composition of centromeric nucleosome is quite controversial. The lines of studies have proposed contradictory models about the accurate structure of centromeric nucleosome. Nevertheless, even though all of the models are quite inconsistent, there is a common consensus in that Cse4 is found in the nucleosome together with H4 histone. Their co-localization is based on the lines of observations showing the physical interaction between Cse4 and H4. Moreover, it was shown that both of them are necessary for the integrity of centromeric chromatin (Glowczewski et al. 2000; Meluh et al. 1998; Black et al. 2004). The interaction of Cse4 and H4 for centromeric nucleosome is further supported by the crystal structure of CENP-A/H4 complex, where CENP-A forms a stable complex with H4 via its α2 helix and α3 helix domain (Sekulic et al. 2010). Although the association of Cse4 with H4 has been demonstrated, whether they associate as dimers or as a monomer and which other proteins are involved in the centromeric nucleosome remains still elusive. Among the many models proposed, there are three popular models about the configuration of centromeric nucleosomes (**Figure 14**).

The first model is the octameric nucleosome model containing Cse4 instead of H3 (Camahort et al. 2009; Kingston et al. 2011; Foltz et al. 2006). Several evidences in various organisms support this octameric configuration, where the centromeric nucleosome is wrapped around an octamer of H2A, H2B, H4 and Cse4 histones with a conventional left-handed twist (Sekulic et al. 2010). This model is based on the high similarities between H3 and Cse4 histones in the structure of stable tetramers that they form with H4 (Luger et al. 1997; Camahort et al. 2009; Sekulic et al. 2010). The studies performed using

chromatin immunoprecipitation (ChIP) assay were able to demonstrate the presence of H2A, H2B and H4 histones together with Cse4 in centromeric nucleosomes. In addition to CHIP assay showing the direct physical interaction of Cse4 with the other three histones, *in vitro* reassembled Cse4 containing octamer confirms the octameric nucleosome arrangement (Camahort et al. 2009). Moreover, this model is supported by the crystal structure of CenH3/H4 tetramer demonstrating that CenH3/H4 tetramer acts as a core of the octameric nucleosome (Sekulic et al. 2010). Finally, biochemical and biophysical analysis indicated that the octameric confirmation containing two copies of Cse4 is the physiologically relevant confirmation (Kingston et al. 2011).

Second model proposes a hexameric nucleosome configuration involving the budding yeast kinetochore protein Scm3, where dimers of Scm3 assemble and maintain the Cse4/H4 tetramer at centromeres (Mizuguchi et al. 2007). Scm3 has been shown to facilitate exclusion of H2A-H2B dimers in preassembled octamer containing Cse4 and consistently H2A/H2B histones were diminished from centromeric DNA, confirming the hexameric configuration. However, this model has been sharply challenged by *in vitro* reconstitution assay where the octameric nucleosome (Cse4:H4:H2A:H2B) is assembled in the absence of Scm3 regions (Camahort et al. 2009). Later, structural analyzes and *in vitro* studies showed that Scm3 disassociates from the octameric complex to stimulate Cse4/H4 tetramer formation and DNA binding (Cho & Harrison 2011). Therefore, these results strengthen the idea that Scm3 is not a component of the centromeric nucleosome, but it rather functions as a Cse4-specific nucleosome assembly factor (Dechassa et al. 2011). Nonetheless, the group defending hexameric nucleosome confirmation raised reasonable arguments against the appeals, as most of the analyzes were based on *in vitro* reconstitution assays, where 601 high affinity nucleosome positioning sequence was used instead of AT-rich CEN sequences (Lowary & Widom 1989). For that reason, *in vitro* reconstitution assay was consequently performed with both 601 and CEN sequences (Xiao et al. 2011). The results have shown that Cse4/H4 tetramer together with Scm3 can be reassembled on CEN DNA. Therefore, they proposed that there are two populations of Cse4 nucleosomes in yeast chromatin; one is octameric and found in non-centromeric locations, whereas second population, found at centromeres, predominantly contains Scm3 and Cse4/H4 tetramer but lacks H2A/H2B. In the proposed hexameric nucleosome model, Cse4/H4 tetramer is located at CDEII element in concord

with previously published results (Camahort et al. 2009; Keith & Fitzgerald-Hayes 2000) and Scm3 dimer binds to periphery of CDEII to stabilize the tetramer. CDEII wrapping around the tetramer brings CBF1 and CBF3 proteins in contact (Kingston et al. 2011; Espelin et al. 1997) which further stabilizes the centromeric nucleosome (Xiao et al. 2011).

The third model suggests a hemisome configuration of centromeric nucleosome, which contains a single molecule of each Cse4, H4, H2A and H2B histones (Dalal et al. 2007). The studies in *D. melanogaster* and mammalian cells have led to the hypothesis that hemisomes wrap around centromeric DNA (Dalal et al. 2007; Dimitriadis et al. 2010). Atomic microscopy demonstrated that in *D. melanogaster* centromeric nucleosome containing

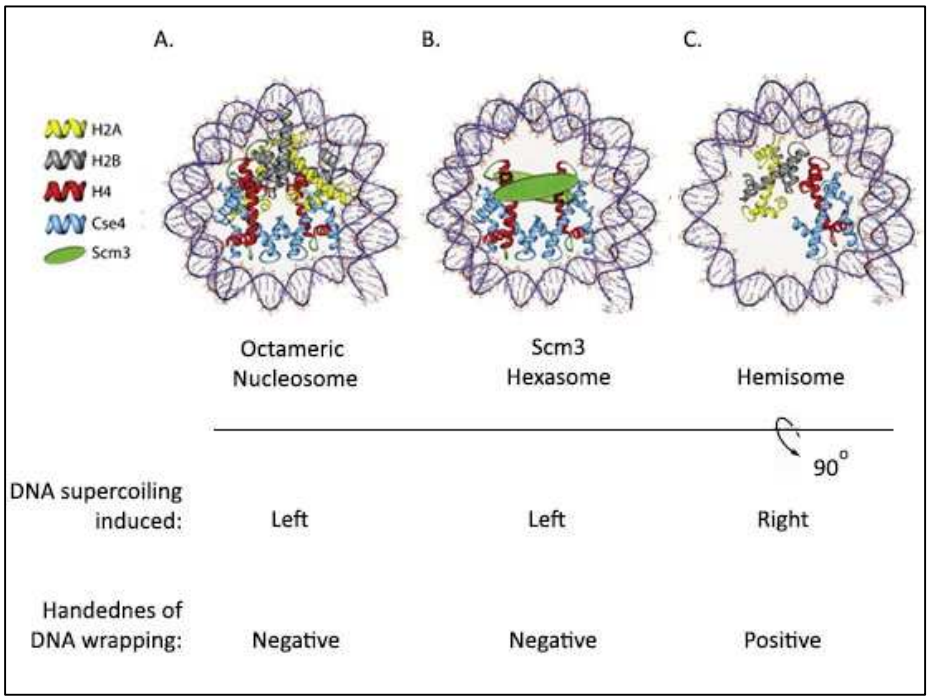


Figure 14: Three most popular models of budding yeast centromeric nucleosome organization. A) The octamer model proposes that centremoric nusleosome is assembled into a canonical-like octameric configuration, containing Cse4. B) The hexasome model involves two molecules of Scm3, where dimers of Scm3 assemble and maintain the Cse4/H4 tetramer at centromeres. C) The hemisome model predicts that centromeric nucleosome contains a single molecule of of each Cse4, H2A, H2B, and H4. Adapted from (Camahort et al. 2009)).

CID is half height of the canonical nucleosome (Dalal et al. 2007). Furthermore, the nuclease protection assay and electron microscopy revealed that CID contains nucleosome wraps around <120 bp of DNA, which is shorter than a

canonical octamer. The smaller size and dimension, as well as the cross-linking assay with CID, favored the hypothesis of hemisomes as a key component of centromeric nucleosomes in *D. melanogaster* (Dalal et al. 2007). The hemisome model was recently broadened to budding yeast (Furuyama & Henikoff 2009). The main claim for the hemisome model in budding yeast comes from the tightly bound proteins on either side of centromeric nucleosomes, i.e. CBF1 and CBF3. CBF1 and CBF3 occupy CDEI and CDEIII elements, leaving only 80 bp CDEII element available for full occupation of centromeric nucleosome. Such a limited region could only permit hemisome occupation (Dalal et al. 2007). Moreover, budding yeast minichromosome system and reconstitution experiments in *D. melanogaster* have provided evidence that centromeric nucleosomes induce positive supercoils, in other words, they wrap centromeric DNA in a right-handed manner (Furuyama & Henikoff 2009). The finding of the instability of left-handed Cse4 containing octamers further suggested the right-handed model of the centromeric nucleosome as the physiologically relevant one (Dechassa et al. 2011). Given that Cse4 containing nucleosomes induce positive supercoils, it seems implausible that centromeric nucleosomes exist as octamers. Finally, the centromeric nucleosome of budding yeast has been characterized at single base-pair resolution using Cse4 CHIP followed by deep sequencing. The resulting nucleosome map revealed that centromeric nucleosome is centered over CDEII element flanked by sub-nucleosomal particles over CDEI and CDEIII elements (Krassovsky et al. 2012). However, the findings showing the dimerization of CenH3 (Zhang et al. 2012; Kingston et al. 2011; Sekulic et al. 2010) overrides the hemisome model. In addition, it was proposed that the small size and dimension of the centromeric nucleosome might have other explanations (Kingston et al. 2011; Black & Cleveland 2011). CBF1 and CBF3 might bind to centromeric DNA on the outside of the nucleosome (Cole et al. 2011) which still enables the histone octamer wrapping. Furthermore, the reduced DNA protection from MNase might not be due to the smaller size of centromeric nucleosome; but rather due the weaker protein-DNA interactions between centromeric histones and DNA (Kingston et al. 2011). Therefore, weaker interactions might allow to higher access to the nuclease.

As deduced from previous paragraphs, despite the tremendous amount of research, there is not yet a general agreement in the centromeric nucleosome structure and content. The discrepancies within the studies might reflect the differences among organisms or cell-cycle stages (Black & Cleveland 2011).

There is also technical part that might lead to different observation for the various models altering the nucleosome structure. For instance, N- or C-terminal tags, fixatives like formaldehyde and *in vitro* assembly conditions might alter or bias the nucleosome structure (Bui et al. 2012). Moreover, the unavailability of an anti-Cse4 antibody, the high sensitivity of centromeric nucleosome to MNase and the scarcity of the centromeric nucleosomes in the whole genome challenge the consensus in the centromeric nucleosome structure.

7.4 Cell Cycle Dependent Fluctuation of Centromeric Nucleosome

In every species, the incorporation of CenH3 at centromeres is cell cycle regulated; however, the exact timing of the deposition might vary among species. In budding yeast, unlike other organisms, Cs4 is incorporated at the S phase (Pearson et al. 2004). The cell cycle regulated deposition of Cse4 and different proposed models for centromeric nucleosome configuration prompted to suggest intermediate forms of centromeric nucleosome that oscillate throughout cell cycle. Quantification of EGFP-tagged Cse4 by fluorescence correlation spectroscopy (FCS) demonstrated that there are two copies of Cse4 per centromere at anaphase and a single copy at the rest of the cell cycle (Shivaraju et al. 2012). These results are further supported by fluorescence resonance energy transfer (FRET) and CHIP assays showing that Cse4-Cse4 interaction is restricted to anaphase (Shivaraju et al. 2012). Notably, both structures contain H2A. The changes in the number of the Cse4 suggested a cell cycle-dependent fluctuation between hemisome, from G1 to mitosis, and octasome during anaphase (Shivaraju et al. 2012; Aravamudhan et al. 2013). This cell cycle coupled oscillation has been recently observed in human cells as well (Bui et al. 2013). Furthermore, it has been demonstrated that the transition from hemisome to octasome coincides with the decrease in the level of Scm3 which might influence the structural transition (Bui et al. 2013; Shivaraju et al. 2012).

OBJECTIVES

OBJECTIVES

The objectives of this thesis are:

- To analyze genome-wide nucleosome positioning in *S. cerevisiae*
- To decipher the nucleosome positioning determinants, in particular the contribution of *cis*-acting factors
- To investigate cell cycle-dependent chromatin dynamics
- To explore the interplay between nucleosome organization and transcription
- To study cell cycle-dependent nucleosome dynamics at replication origins and its implication in replication origin activity
- To characterize centromeric nucleosome composition and positioning

MATERIALS & METHODS

1 Materials

1.1 Chemicals and Consumables

Chemicals	Provider
Acetic Acid	Sigma Aldrich
Acrylamide/Bis-Acrilamide 30% Solution, 29:1	Sigma Aldrich
Agar Bacto	BD
Agarose	Sigma Aldrich
Albumin Bovine Serum, BSA	Sigma Aldrich
Ammonium Persulfate	Sigma Aldrich
Bromophenol Blue	Sigma Aldrich
Calcium Chloride	Panreac
Chloroform	Sigma Aldrich
dNTP Mix	Roche
dNTP Set	Invitrogen
Dimethyl Sulfoxide, DMSO	Sigma Aldrich
DTT, Dl-Dithiothreitol	Sigma Aldrich
Ethylenediaminetetraacetate-Disodium, EDTA	Sigma Aldrich
Ethanol Absolute	Panreac
Ethidium Bromide	Sigma Aldrich
Alpha-factor Mating Pheromone	Genscript
Ficoll	Fisher Bioreagents
Formaldehyde Solution	Sigma Aldrich
Gel Loading dye (6x)	N.E. Biolabs
Glucose D-(+)	Sigma Aldrich
Glycerol	Acros Organics
Glycine	Sigma Aldrich
Hoechst	Invitrogen
Isopropyl Alcohol. 2-Propanol	Sigma Aldrich
Magnesium Chloride	Sigma Aldrich
B-Mercaptoethanol	Sigma Aldrich
Methanol	Panreac
Mini-Protean Tgx Gels, 4-20%	Bio-Rad
Nonidet® P40	Fluka
PBS 10x (Phosph. Buffered Salt), ph 7.4	Sigma Aldrich
Peptone Bacto	BD
Phenol Solution	Sigma Aldrich
Phenol:Chloroform:Isoamyl Alcohol 25:24:1	Sigma Aldrich
Phenymethanesulfonyl Fluoride, PMSF	Sigma Aldrich
Pipes, Disodium Salt	Sigma Aldrich
Potassium Acetate	Sigma Aldrich
Potassium Chloride	Sigma Aldrich
Potassium Dihydrogen Phosphate	Panreac
Potassium Hydroxide	Sigma Aldrich
Propidium Iodide Solution, PI	Sigma Aldrich

Sodium Chloride	Sigma Aldrich
Sodium Docecyl Sulphate, SDS	Sigma Aldrich
Sodium Hydroxide	Fluka
Sorbitol, D	Sigma Aldrich
Spermidine	Sigma Aldrich
Sucrose, Saccharose	Sigma Aldrich
Sybr Safe DNA Gel Stain	Invitrogen
TBE(10x)Tris-Borate-Edta Buffer	Sigma Aldrich
Temed, N,N,N ,N -Tetramethylethylenediamine	Sigma Aldrich
Triton® X-100	Sigma Aldrich
Trizma® Base	Sigma Aldrich
Yeast Extract Bacto^{tm} BD	BD
1 kb DNA ladder (1 µg/µl)	Invitrogen
100 bp DNA ladder (1 µg/µl)	Invitrogen

1.2 Enzymes

Enzymes	**Provider**
Microccocal Nuclease	Sigma Aldrich
Taq DNA polymerase recombinant	Invitrogen
Pronase	Sigma Aldrich
Proteinase K	Roche
RNase A	Roche
Zymolase 20T	Seikagaku
Zymolase Long Life	Genotech

1.3 Kits

Kits	**Provider**
NucleoSpin® Gel and PCR Clean-up	Macherey-Nagel
Pierce Silver Stain Kit for Mass Spectrometry	Thermo Scientific
Qubit® dsDNA BR Assay Kit	Invitrogen
Qubit® dsDNA HS Assay Kit	Invitrogen
Qubit® dsDNA Protein Assay Kit	Invitrogen
Qubit® RNA Assay Kit	Invitrogen
RNeasy® Mini kit (50)	Qiagen
ECL Prime Western Blotting Detection Reagent	Life Sciences

1.4 Oligonucleotides

Oligonucleotides	**Sequence (5’ - 3’)**
5´CEN14	GGT TCT AGT TAG TCA CGT GCA GCT TTT TAA
3´CEN14	GCT AGG CGC CTA AAC GCA GAT ATC CTT AAA
5´CEN4	GTC ACA TGC TTA TAA TCA ACT TTT TTA AAA ATT
3´CEN4	GTT TTA TGT TTC GGT AAT CAT AAA CAA TAA ATA
5´CEN6	CAT CAC GTG CTA TAA AAA TAA TTA TAA TTT AAA
3´CEN6	TTT ACA TCT TCG GAA AAC AAA AAC TAT TTT TTC

1.5 Antibodies

Name	Description/ Dilution/ Supplier
α-H3-N (Rabbit)	Polyclonal/ 1:8000 (WB), 5 µl (ChIP)/ Sigma H9289
α-H3K4ac (Rabbit)	Polyclonal/ 1:8000 (WB), 5 µl (ChIP)/ Sigma SAB4800022
α-IgG (Rabbit)	Secondary, goat (HRP-conjugated)/ 1:5000/ Abcam AB6721

1.6 Instruments

Instruments	Provider
Balance of precision	Control Tecnica
Balance	Rubilabor
Beckman Coulter *EPICS® XL* flow cytometer	Beckman
Bioruptor system	Diagenode
Centrifuge 5415 R	Hucoa
Centrifuge 5804R	Hucoa
Centrifuge Minispin Plus	Eppendorf
Centrifuge Unifuge, Microcentrifuge	Fisher Scientific
Image Quant RT	GE Healthcare
Incubator 30ºC Model B12	Thermo Scientific
Incubator Innova® 42/42R Shaker	New Brunswick
Magnetic stirrer	Fisher Scientific
Microscope DM300	Leica
Milli-Q Ultrapure Water System	Millipore
Mini-Gel Caster - 170-4422	Bio-Rad
Mini-PROTEAN Tetra Cell system	Bio-Rad
NanoDrop Spectrophotometer 2000	Thermo Scientific
Nikon E600 fluorescence microscope	Nikon
pH-meter GLP21, temperature sensor	Crison
Plataform Rocking Shaker	Heidolph
Peristaltic Pump	GE Healthcare
Sonicador (VC750)	Vibracell
Qubit, Fluorometer	Invitrogen
Thermocycler, PCR	Eppendorf
Thermostatic Water Bath	Thermo Scientific
Concentrator 5301	Beckman
Vacuum Concentrator Speeddry	Eppendorf
Vortex B15012	Fisher Scientific

1.7 Yeast strains

Two strains have been used throughout this work.

BY4741 strain (*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0*) derived from S288C was purchased from Life Technologies.

PPY1 strain (*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 bar1::leu2)* was taken from Oscar Aparicio's lab at the University of Southern California, USA. PPY1 strain was derived from BY4741, transformed with cut pZV77 and confirmed with bar assay.

The yeast strains were routinely grown at 30 °C on YPD-agar plates or in liquid YPD, shaking at 200 rpm.

YPD growth medium
1% Bacto-yeast extract
2% Bacto Peptone
2% D-(+)-Glucose (Dextrose)
2% Agar (only for plates)

2 Methods

2.1 DNA Specific Methods

2.1.1 Genomic DNA Preparation

Cultures of *S.cerevisiae* strain BY4741 were grown in 30-40 ml of YPD rich medium for 20 h shaking at 30 °C. The cells were harvested by centrifugation for 5' at 4500 g (5000 rpm) at RT using a bench top centrifuge (Eppendorf 5804R) and resuspended in 10 ml Buffer Z1. 1 ml of 10 mg/ml zymolase (Seigaku, Inc.) was added and incubated at 30 °C for about 30', until 90% of the cell walls were lysed. The so-called spheroplasts were pelleted by 5' centrifugation at 4500 g (5000 rpm) and then incubated at 65 °C for 30' with 5 ml Tris-EDTA (TE) buffer containing 1% (w/V) SDS. Subsequently, 2 ml of 5 M KOAc was added and incubated for 1 h on ice. The pellet was precipitated by centrifugation for 10' at 16100 g (13200 rpm). The supernatant was transferred into new microtubes containing 1 volume (V) of phenol and centrifuged for 5' at 16100 g (13200 rpm). The upper layer was transferred into new microtubes containing 3 V of EtOH and 1/10 V of 3 M NaOAc and incubated for 1 h at -20 °C. The pellet was recovered by centrifugation for 20' at 16100 g (13200 rpm) at 4°C, washed with 500 µL 70% EtOH, lyophilized and resuspended in 20-50 µL TE buffer. RNase A was added to a final concentration of 0.5 mg/ml and incubated for 1 h at 37 °C. DNA was precipitated with 3 V of EtOH, 1 /10 V of NaOAc for 20' at 16100 g (13200

rpm) at 4°C. The pellet was washed with 500 µL 70% EtOH, lyophilized and resuspended in 20-50 µL TE.

Buffer Z1 (for Zymolase digestion 1)	TE buffer
1 M sorbitol	50 mM Tris, pH 7.4
20 Mm Tris-Cl, pH 7.4	20 mM EDTA
15 µM β-mercaptoethanol	

2.1.2 Digestion of Genomic DNA

Genomic DNA samples were fragmented either by MNase digestion or *Bioruptor* disruption. For MNase digestion, 100 µl aliquots with 2 µg of purified genomic DNA were transferred to Eppendorf tubes. $CaCl_2$ solution was added to a final concentration of 3 mM. The samples were digested with micrococcal nuclease (MNase; Sigma-Aldrich, Inc.) at concentrations of 0, 0.01, 0.03, 0.06 and 0.1 U, respectively, in a water-bath at 37 ^{o}C for 10'. The digestion reactions were quenched with 2 µl of 500 mM EDTA and purified by ethanol precipitation. Fragmentation by *Bioruptor* system was performed with 2 µg of DNA sonicated during 0, 5, 10 and 15' (at intervals of 10 s on-30 s off), respectively. In both approaches, the purified samples were examined by 2% (w/V) agarose gels supplemented with SybrSafe and reactions containing a fragment size of 100-350 bp were selected for DNA sequencing.

2.1.3 Nucleosomal DNA Preparation

5 ml of BY4741 culture was grown overnight in YPD rich medium and diluted to an OD_{600} of 0.2 using 100 ml of fresh YPD media next day. The culture was further grown at 30 ^{o}C until an OD_{600} of 0.8–0.9. Cells were cross-linked with the 2.8 ml of 37% (V/V) formaldehyde for 15' on the orbital shaker at RT and the reaction was stopped by the addition of 5 ml of 2.5 M glycine. Cells were harvested by centrifugation at 1620 g (3000 rpm, 5804R Eppendorf centrifuge) for 5' at 4 ^{o}C, washed with ice-cold 20 ml PBS buffer and resuspended in 15 ml buffer Z2. Subsequently, 750 µl of 10 mg/ml zymolase was added and incubated at 30 ^{o}C for about 30' until 90% of the cell walls were lysed. The spheroplasts were pelleted by centrifugation for 5' at 4500 g (5000 rpm) and washed twice with 5 ml of buffer Z2. The pellet was resuspended in 4 ml buffer Y and transferred into new eppendorf tubes with 500 µl of sample each. Different MNase digestion reactions were setup at concentrations of 0, 0.04, 0.08, 0.16 and 0.12 U, respectively. The digestion reactions were incubated at 37 ^{o}C for 30' and stopped by adding 20 µl of 500

mM EDTA. The samples were first treated with 25 µl of 10 mg/ml DNase-free RNase for 1 h at 37 °C, then with 25 µl of 10 mg/ml proteinase K for 2 h at 37 °C, and incubated overnight at 65 °C to reverse the cross-link. Subsequently, 1 V of phenol was added to the samples, vortexed, spun down for 5' at 13000 g (16000 rpm) and precipitated with 3 V of 100% EtOH and 1/10 V of NaOAc by centrifugation for 20' at 13200 g (16100 rpm) at 4 °C. The pellet was washed once with 70% EtOH, lyophilized and resuspended in 20-50 µl dH2O. DNA fragments were examined by 2% agarose gels supplemented with SyberSafe. Those reactions, containing at least 90% mononucleosomal DNA fragments, were selected for sequencing.

In addition, MNase over- and under-digested samples were included. Over-digested samples were obtained at a 0.12 U MNase concentration, which yielded only mononucleosomes. Under-digested samples were obtained at a 0.04 U, yielding mono-, di- and tri-nucleosomes.

Buffer Z2 (for Zymolase digestion 1)
1 M sorbitol
50 Mm Tris-Cl, pH 7.4
10 µM β-mercaptoethanol

Buffer Y
10 mM Tris, pH 7.4
50 mM NaCl
1 mM $CaCl_2$
5 mM $MgCl_2$
0.075% (V/V) Nonidet P40
1 mM β-mercaptoethanol
500 µM spermidine

2.1.4 Determination of DNA concentration

DNA concentration of purified samples was determined by UV absorption measurement at 260 nm using *Nanodrop* spectrophotometer (Thermo Scientific, Inc.) and by fluorescence emission using *Qubit* fluorometer (Invitrogen, Inc.)

2.1.5 Agarose Gel Electrophoresis

DNA fragments from 50 to up 10 000 bp were separated and analyzed on 1.5-2% agarose gels in 1 × TBE and 0.5 µg/ml EtBr or Syber Safe using horizontal electrophoresis systems. The DNA samples were mixed with 1/5 V of 6 X Loading Buffer. The gels were run at 80-120V for 30' to 1 h at RT. The gels were analyzed using Image Quant RT ECL Imager (GE healthcare).

TBE buffer
90 mM boric acid
90 Mm Tris-Cl, pH 8
20 mM EDTA

6 X Loading Buffer Y
50 mM Tris-HCl, pH 7.6
0.05% (w/v) bromophenol blue
0.05% (w/v) xylene cyanol
50% (v/v) glycerol
10 mM EDTA

2.1.6 Analysis of DNA with Bioanalyzer

The integrity and size distributions of DNA fragments within a purified sample were determined using the microfluidics-based platform *Bioanalyzer* (Agilent), performed at the Scientific-Technology Services from the University of Barcelona.

2.2 Chromatin Analysis along Cell Cycle

2.2.1 Cell Cycle Synchronization

100 ml culture of BY4741 or PPY1 strain was grown using fresh YPD media until an OD_{600} of 0.2. Then, alpha-factor mating pheromone (GenScript) was added to a final concentration of 10 µM for BY4741 (100 nM for pPY1 strain) to allow cell synchronization in late G1 for 2 h, till 100% of the cells were un-budded and formed shmoos. The alpha-factor was removed by harvesting the cells with centrifugation for 10' at 4500 g (5000 rpm). The arrested cells were inoculated in 100 ml fresh YPD rich medium. 25 ml samples were collected immediately after release from G1 arrest at different intervals for 90'. The intervals varied from 3 to 15' depending on the experiments, which are specified in the results section.

2.2.2 Monitoring Cell Cycle Synchrony

Cell synchrony was monitored by three approaches: flow cytometry (FACS), fluorescence microscopy and budding index calculation. 50-500 µl aliquots were taken from each.

2.2.2.1 FACS Analysis

For FACS analysis, 50 µl of cells were fixed with 1 ml of 100% EtOH for 1 h at 4 ^{o}C, spun down for 5' at 13200 g (16100 rpm) and washed once with 500 µl of 1 X SSC buffer. The pellet was resuspended in 500 µl of 1x SSC buffer. 25 µl of 10 mg/ml RNase A was added and the samples were incubated for 1 h at 50

°C. Later, the samples were treated with 25 µl of 10 mg/ml Proteinase K for 2 h at 50 °C, briefly sonicated by using the *Bioruptor* system for 5' (at intervals of 10 s on-30 s off) and mixed with 500 µl SSC buffer containing 0.1 mg/ml propidium iodide (PI, Sigma-Aldrich). Fluorescence emitted from DNA-intercalated PI was measured by Beckman Coulter *EPICS® XL* flow cytometer.

SCC Buffer (20x)
150 mM NaCl
15 mM sodium citrate
pH 7.8

2.2.2.2 Budding Index Calculation and Fluorescence Microscopy

Cell cycle phase was also monitored by fluorescence microscopy and budding index calculation. For these purposes, 500 µl cells were briefly sonicated by using the *Bioruptor* system for 5' (at intervals of 10 s on-30 s off), washed with 500 µl 1x PBS and fixed with 70% EtOH for 1 h at 4 °C. Fixed cells were then resuspended in 200 µl PBS containing Hoechst stain (30 µg/ml). Finally, cells were placed on a glass slide and visualized by fluorescence microscopy (*Nikon* E600 microscope). For budding index calculation, a sample from EtOH fixed cells was placed on a hemocytometer and visualized under a student microscope (*Leica* DM300) to count the number of budded and un-budded cells.

2.2.3 Nucleosomal DNA Preparation

The selected samples after microscopy and FACS analyzes were fixed immediately with 700 µl of 37% formaldehyde and the nucleosomal DNA preparation was performed as described in 2.1.3. Different digestion reactions were setup with MNase at 0.1, 0.15 and 0.2 U concentrations, respectively. DNA resulting fragments were examined on 2% agarose gels. Those reactions containing at least 90% mononucleosomal DNA fragments were selected for sequencing.

2.2.4 RNA Isolation

For gene expression analysis, 8 ml of samples were collected at the same intervals as nucleosomal DNA samples in 15 ml falcon tubes, which were filled with icy-water. The samples were harvested at 6000 rpm for 3-4', frozen in liquid nitrogen and stored at - 80°C. Total cellular RNA was extracted using the *RNeasy* kit (Qiagen), following the manufacturer's instructions after the

spheroplasting protocol (0.5 mg/ml zymolase). The total RNA was hybridized to Affymetrix GeneChip Yeast Genome 2.0 arrays for gene expression analysis.

2.2.5 Determination of RNA Concentration

RNA samples were quantified by UV absorption measurement at 260 nm using *Nanodrop* spectrophotometer (Thermo Scientific, Inc.) and fluorescence emission using *Qubit* fluorometer (Invitrogen, Inc.)

2.2.6 Gene Expression arrays

The total RNA sample content was hybridized to Affymetrix GeneChip Yeast Genome 2.0 arrays for gene expression analysis. The resulting raw *CEL* files were imported and processed with *R/Bioconductor* framework (Gentleman *et al.,* 2004). Quartile normalization was applied, expression ratios from the array were converted to log2 values and the quality was assessed using the *MDA* package (http://www.bioconductor.org).

2.3 Centromeric Nucleosome Characterization

2.3.1 Nuclei Isolation

For centromeric nucleosome isolation, chromatin was treated differently. Nuclei isolation protocol was adapted from Wu *et al.,* 2000. 5 ml of yeast culture was grown overnight and diluted the next day to an OD_{600} of 0.2 using 200 ml fresh YPD media. The culture was further grown at 30 ^{o}C until reaching an OD_{600} of 0.8–0.9. For cell cycle analyzes, cultures were arrested at late G1 by alpha-factor mating pheromone (GenScript). Samples collected at different time intervals were cross-linked with 2% (V/V) formaldehyde for 5’ while shaking at 30 ^{o}C and quenched by the addition of 125 mM glycine. Otherwise, formaldehyde fixation procedure was skipped. Cell cultures were harvested at 1000 g (2357 rpm) for 10’ at RT using a bench-top centrifuge. Supernatants were discarded and the pellets were resuspended in 20 ml of 100 mM PIPES, pH 9.4 containing 10 mM DTT and incubated for 10’ at 30 ^{o}C with shaking. Resuspended cells were centrifuged at 1000 g for 10’ at RT, resuspended in 10 ml of YEP 0.2% glucose buffer. 30 U/ml zymolase (Genotech) was added and incubated at 30 ^{o}C shaking for 30’ until 90% of the cells were lysed. Cells were harvested at 1000 g for 10’ at RT, resuspended in 20 ml of YEP 1% glucose buffer and incubated for 20’ at 30

°C. Resuspended cells were spun down at 1000 g for 10' at RT, washed twice with 5 ml of permeabilization buffer and resuspended in 1-3 ml of buffer P. Nuclei were diluted 200X and counted on a hemocytometer. Nuclei aliquots were frozen in liquid nitrogen and stored at -80 °C.

YEP 0,2% glucose buffer
1% bacto-yeast extract
1% bacto peptone
0.5% NaCl (w/V)
50 mM KH_2PO_4
0.2% glucose (w/V)
0.6 M sorbitol

YEP 1% glucose buffer
1% bacto-yeast extract
1% bacto peptone
0.5% NaCl (w/V)
1% glucose (w/V)
0, 7 M sorbitol

Permeabilization Buffer
20 mM PIPES-KOH pH. 6.8
150 mM CH_3CO_2K
2 mM $Mg(C_2H_3O_2)_2$
0,4 M sorbitol
10% DMSO

2.3.2 Nuclei Digestion

To ensure nucleosomal DNA fragments of a desired length, we performed a digestion optimization at different MNase concentrations with a small amount of nuclei from every batch preparation. One nuclei aliquot was thawed on ice, pelleted at 3000 g for 5' with a table-top centrifuge and resuspended in EcorI buffer with 3 mM $CaCl_2$ or in buffer Y, in a final volume of 80 µl. Different digestion reactions were setup with MNase concentrations of 0, 0.05, 0.1 and 0.15 U. The digestion reactions were incubated at 37 °C for 30' and stopped by adding 20 µl of stop solution.

To check fragment lengths, small sample aliquots were taken and treated with DNase-free RNase (1 mg/ml) for 1 h at 37 °C, to degrade all RNA content. In case the samples had been previously crosslinked, 0.8 mg/ml proteinase K was added and incubated overnight at 65 °C to reverse cross-linking and remove protein content. Otherwise, 1 M NaCl and 1% SDS at final concentration were directly added and samples were vortexed vigorously. After protein removal, either by proteinase K or NaCl/SDS, DNA was purified by phenol extraction and ethanol precipitation, as described previously. Purified DNA was dissolved in distilled water and examined by 2% agarose

gels. DNA samples containing the optimal fragment size were selected for sucrose gradient centrifugation.

ECORI buffer	Stop Solution
S50 mM Tris-HCl (pH 7.5)	10 mM EDTA
10 mM $MgCl_2$	1% SDS
100 mM NaCl	
0.02% Triton X-100	
0.1 mg/mL BSA	

2.3.3 Chromatin Immunoprecipitation (ChIP)

In order to enrich our samples with the centromeric nucleosomes in our samples, ChIP experiments with anti-Histone H3 antibodies were performed. Cells were prepared and spheroplasted as described in 2.3.1. Then, spheroplasts were washed twice and resuspended in SPC buffer with with 1 mM PMSF and 1× LPC (l0 µg/ml each of leupeptin, pepstatin A, and chymostatin). Nuclei were digested by MNase in EcorI buffer, as described in 2.3.2 and pelleted 3,000 rpm for 10 min at 4°C. The supernatant (S1) was recovered and Triton X-100 was added to 0.1% before immunoprecipitation. The pellet was resuspended in TE buffer with 0.1% Triton X-100, 1× LPC, and 1 mM PMSF, and rotated at 4°C overnight to further extract chromatin. The sample was pelleted at 8,000 rpm for 10 min at 4°C the second supernatant (S2) was combined with S1 and PBS350 was added to the S1 to a final salt concentration of 100 mM.

For immunoprecipitation, the soluble chromatin was mixed with the appropriate antibody: either 5 µl of anti-H3K4ac (control; Sigma, SAB4800022) or 5 µl of anti-H3-N (Sigma, H92899) for overnight at 4°C. Before use, Dynabeads Protein G (Life Technologies, 70 µl per ChIP) were washed in lysis buffer containing 4 mg/ml bovine serum albumin. The beads were added to the chromatin and incubated at 4°C for 1 h. Beads were pelleted and the supernatant was saved as unbound material for subsequent sucrose gradient centrifugation (2.3.4). An aliquot of the unbound material was taken for DNA and protein analysis. Protein content was precipitated by 10 mM $MgCl_2$ for 1 h at 13200 g (16100 rpm) at 4 ^{o}C and dissolved in Laemmli Buffer (2.4.1). The IP-bound material was washed twice with lysis buffer, twice with lysis buffer containing 500 mM NaCl, twice in wash buffer and once in TE. To elute DNA, half of the beads were dissolved in 50 mM Tris, pH 8, 10 mM EDTA and 1% SDS and were incubated for 10 min at 65 ^{o}C. DNA

from bound and unbound chromatin was isolated with proteinase K treatment followed by phenol extraction, ethanol precipitation. Half of the beads were dissolved in Laemmli Buffer and prepared for protein content analysis (Materials&Methods 2.4).

SPC Buffer
1 M Sorbitol
20 mM Pipes, pH 6.3
0.1 mM $CaCl_2$

PBS350
PBS, pH 8.0
350 mM NaCl
1x LPC
1 mM PMSF
1 mM EDTA

Lysis Buffer
50 mM HEPES-KOH, pH 7.5
140 mM NaCl
1% Triton X-100

Wash Buffer
10 mM Tris-HCl pH 8.0
250 mM LiCl
0.5% Nonidet-P40
0.5% Na-deoxycho**late**
1 mM EDTA

2.3.4 Sucrose Gradient Centrifugation

Nuclei digested samples were separated according to their density by 5-15% sucrose gradient centrifugations. 12 ml of 5% and 15% of sucrose in gradient buffer were respectively dispensed into two separate chambers in a linear gradient maker. Using a peristaltic pump, a 5-15% linear sucrose gradient was prepared in Beckman polycarbonate tubes and stored for at least 30' at 4 ^{o}C. 1-5 µg of MNase digested chromatin was layered on top of the gradient and centrifuged on a Beckman SW41 rotor at 30.000 rpm for 16 h at 4^{0}C. Fractions of 500 µl-1 ml were collected in 1.5 ml Eppendorf tubes and stored at 4 ^{o}C.

DNA fragment sizes were inspected by taking a small aliquot from each fraction and isolating its DNA with phenol extraction and ethanol precipitation as described in the previous section.

Gradient Buffer
100 mM NaCl
10 mM Tris
0.2 mM EDTA
100 µg/ml PMSF

2.3.5 Centromeric DNA Identification by Polymerase Chain Reaction

To identify those fractions containing centromeric nucleosomes, polymerase chain reaction (PCR) amplification was employed using sequence-specific oligonucleotide primers for centromeric DNA sites, as listed in section 1.5. To this end, PCR reactions in a final volume of 24 µl were prepared containing 200 nM of forward and reverse primers, 1.5 mM $MgCl_2$, 200 µM of dNTP mix, 1X of PCR mix and 2.5 U Taq polymerase (Invitrogen). 1 µl of each fraction was added to the PCR mixture and reactions were run on a Eppendorf thermocycler using the protocol indicated in the table below.

Step	Temperature (^oC)	Time
1	95	5’
2	95	30 s
3	58	30 s
4	72	30 s
5	Go to step 2, 32 times	
6	72	5’
7	4	forever

The PCR products were mixed with 6x gel loading buffer (N.E. Biolabs) and analyzed on a 2% agarose gel.

2.4 Protein Content Analysis

2.4.1 SDS-polyacrylamide Gel Electrophoresis (SDS-PAGE)

One third of each gradient fraction was taken for nucleosome-associated protein content identification. 10 mM $MgCl_2$ was added to each fraction and centrifuged at maximum speed for 1 h at 4 ^oC. The supernatant was discarded carefully and the pellet was dissolved in 1x Laemmli buffer, heat-denatured for 5’ at 95°C and analyzed on a SDS-polyacrylamide gel (8.3× 7.3 cm, 1 cm) using a vertical Mini-Protean-III-Gel-System (BioRad). 12 w/V% SDS-polyacrylamide gels (acrylamide/N,N’-methylene-bis-acrylamide 29:1) were prepared according to the standard protocols, indicated below (Sambrook and Russell, 2001). For better resolution 4-20% Miniprotean gels were used. The heat- denatured samples together with histone control samples were loaded on the gel and run in 1x SDS running buffer at 100 V until the bromophenol blue migration front reached the bottom of the gel.

	Separating Gel (12%, 5 ml)	Stacking Gel (5%, 3 ml)
H_2OMilliQ	1.6	1.7
1.5 M Tris HCl, pH 8.8	1.3 ml	
0.5 M Tris HCl, pH 6.8		0.76
SDS 10%	50 µl	30 µl
Acrylamid/Bis 30% w/V	2	0.5 ml
APS 10%	50 µl	30 µl
TEMED	3 µl	3 µl
Sum	5 ml	3 ml

5x Laemmli buffer
300 mM Tris/HCl pH 6.8
10% SDS
50% glycerine
0.05% Bromphenolblue
5% -Mercaptoethanol

SDS running buffer (10x)
250 mM Tris pH 8.3
1.92 mM Glycine
1% SDS
250 mM Tris pH 8.3

2.4.2 Silver Staining of SDS-PAGE Gels

Upon electrophoresis, SDS-PAGE gels were washed once with distilled water and fixed with a 10% acetic acid and 50% methanol solution, from 1 h to overnight. Silver staining was performed using SilverQuest™ Silver Staining Kit (Invitrogen). Later, stained protein bands were visualized on Image Quant RT ECL.

2.4.3 Mass Spectrometric Identification of Nucleosome-Associated Proteins

After silver stained gel visualization, the bands of interest were cut and sent to Mass Spectrometry core facility, IRB. The protein bands cut from the gel were digested with trypsin at Parc de Scientific (PCB). After the digestion samples were dried in the speed-vac and resuspended in 40 µl 1% formic acid and loaded to a 180 µm × 2 cm C18 Symmetry trap column (Waters) at a flow rate of 15 µl/min using a nanoAcquity Ultra Performance LCTM chromatographic system (Waters Corp., Milford, MA). Peptides were separated using a C18 analytical column (BEH130™ C18 75 mm × 25 cm, 1.7 µm, Waters Corp.) with a 80’ run, comprising three consecutive steps with linear gradients, followed by isocratic and stabilization to initial conditions. The column outlet was directly connected to an Advion TriVersa NanoMate

(Advion) fitted on an LTQ-FT Ultra mass spectrometer (Thermo). The mass spectrometer was operated in a data-dependent acquisition (DDA) mode. The spectrometer was working in positive polarity mode and singly charge state precursors were rejected for fragmentation. Peptides with a q-value lower than 0.1 and a false discovery rate (FDR) < 1% were considered as positive identifications with a high confidence level.

2.4.4 Western Blot Analysis

For validation ChIP control efficiency, protein enrichment in bound material was analyzed by Western Blot. Proteins separated by SDS-PAGE gel were transferred to a polyvinylidene fluoride (PVDF-FL) membrane equilibrated for 30' in methanol. The gel was sandwiched in the following order: Whatman paper, acrylamide gel, PVDF-FL membrane and another layer of Whatman paper, soaked in blotting buffer. Care was taken not to trap air-bubbles between the layers. Sandwiches were placed between two fiber pads and inserted into BioRad Mini Trans-Blot Electrophoretic Transfer Cell Cassette. The cassete was placed into the buffer tank and protein transfer was performed in cold blotting buffer for 90' at 100 V. Membrane was removed and incubated in 5% PBST blocking milk for 1 h at RT. Then, membrane was incubated with primary antibody (anti-H3K4ac or anti-H3-N) diluted at 1:8000 in 5% PBST blocking milk overnight at 4^{o}C. After removal of primary antibody, the membrane was washed three times for 10' with PBST. The membrane was incubated with secondary antibody; diluted 1:5000 in 5% PBST blocking milk for 1 h at RT. Membranes were then washed three times for 10' with PBST. Proteins were detected by rinsing the membranes with a 1:1 dilution of Chemiluminescence Substrate (LifeSciences) that allows visualization of horseradish peroxidase-conjugated secondary antibodies by exposing the membranes to X-ray film (Kodak).

Blotting Buffer	5% PBST blocking milk
1.8% glycine	PBS
0.4% Tris	0.01% Tween (v/v
20% methanol	5% non-fat milk

2.5 Molecular Genetics Methods

DNA tilling microarray was performed by IRB Functional Genomics Core Facility. High-throughput DNA sequencing was performed by the

biotechnology company Fasteris, based in Switzerland. Gene expression arrays were performed by the IRB Functional Genomics core facility. Data processing and DNA physical descriptor algorithm development were performed at Prof. Orozco's computational lab.

2.5.1 DNA Tiling Microarray

MNase-digested nucleosomal and genomic DNA was hybridized to *S. cerevisiae* Tilling 1.0R Array (Affymetrix). The CEL files, which deliver the measured intensities from the scanner, have been merged and probe sequences were mapped using Package Starr (Zacher et al. 2011). The coverage was calculated, the data was processed and noise was removed by using nucleR package (Flores & Orozco 2011) .

2.5.2 High-throughput DNA Sequencing And Read Generation

MNase cleaved DNA samples were single- or paired-end sequenced on an Illumina *HiSeq 2000* or Solexa Sequencer in 54 cycles with 7 extra cycles for multiplex indexing and subsequently pre-processed with a standard Illumina GA base-call pipeline using ELAND 1.5.1 and CASAVA 1.7 software. High-throughput sequencing reads of 38 or 50 bp length (for single-end) or of 50, 54 or 100 bp length (for pair-end) were obtained in *qfasta* format.

2.5.3 Read Genome Alignment and Pre-Processing

Reads were aligned onto the *SacCer3* genome with *Bowtie* aligner (Langmead et al. 2009) allowing a maximum of three mismatches and an insert length of 500 bp. Genome sequences were obtained from the UCSC genome browser (http://hgdownload.cse.ucsc.edu/goldenPath/sacCer1/bigZips/, date of access: 5th February 2010). Due to the presence of repetitive sequences along the genome, those reads that could be ambiguously aligned on multiple regions were mapped to all the possible places, avoiding depleted region artifacts.

2.5.4 Read Import and Duplicate Removal

Sequencing reads were imported using a high-throughput sequencing data analysis library for R/Bioconductor software (http://www.r-project.org, http://www.bioconductor.org/) (Gentleman et al. 2004). Reads from different strands were shifted downstream by using nucleR library (Flores & Orozco 2011) to align the read 5'-end in one strand with the read 3'-end in

the opposite strand. Largely over-represented reads were eliminated with *HtSeqTools* (Planet et al. 2012) to reduce PCR amplification biased artifacts.

2.5.5 Read Genome Coverage Calculation and Nucleosome Calling

Single-end reads were resized to 50bp and shifted downstream to align reads mapping in opposite strands using nucleR (Flores & Orozco 2011). Paired-end reads were trimmed to 50bp maintaining the original center. Genome-wide coverage was normalized using the total number of reads in every experiment and scaled by a factor of 10^6 to obtain the units of reads per million (r.p.m.). Peak calling was performed after noise filtering using nucleR parameters: peak width = 125bp, peak detection threshold = 35%, maximum overlap = 50bp.

2.5.6 Nucleosome Profile Clustering at TSSs

Using the nucleosome calls obtained previously, every gene was classified according to their nucleosome architecture around the TSS. The closest nucleosome at or immediately downstream TSS was annotated as the +1 nucleosome. The nucleosome immediately upstream of the +1 nucleosome was annotated as the -1 nucleosome. After a visual analysis of the classifications, nucleosome calls were considered as well-positioned (W) when nucleR peak width score (score_w; positioning) and height score (score_h; coverage) were higher than 0.4 and 0.6 respectively (even results are quite robust to small perturbations in these values, for a fine analysis of sample variability we considered the numerical differences among all the scores as described in the next section). Otherwise, the nucleosome call was considered fuzzy (F). Accordingly with previous observations (Zaugg & Luscombe 2011), the NFR was defined as the distance between the dyads of the -1/+1 nucleosome and it was annotated as “open” if this distance was greater than 215bp or as “closed” otherwise. The classification of a given gene was determined by the positioning of the -1 nucleosome, the width of the NFR and the positioning of the +1 nucleosome. Special cases such as when the -1 nucleosome was more than 300 bp further from the TSS (annotated as M, missing), the -1/+1 nucleosome calls were overlapped or when the regions -300:+300 bp had more than a 25% of uncovered bases were excluded from the analysis.

2.5.7 Physical Descriptors and Nucleosome Deformation Energy

Parameters describing the equilibrium geometry and deformability of naked DNA were derived from long atomistic MD simulations. In short, we collected equilibrium MD trajectories (150 ns long; T=298 K, P=1 atm.) in water (more than 9,000 TIP3P molecules Na^+ as counter ion) using state of the art simulation protocols for four duplexes, which contain the ten unique dinucleotide steps (steps (d(GG)·d(CC), d(GC)·d(GC), d(GA)·d(YC), d(GT)·d(A·C), d(AG)·d(CT), d(AA)·d(TT), d(AT)·d(AT), d(CG)·d(CG), d(CA)·d(TG) and d(TA)·d(TA)): d((GCCTATAAACGCCTATAA)·d(TTATAGGCGTTTATAGGC), d(CTAGGTGGATGACTCATT)· d(AATGAGTCATCCACCTAG), d(CACGGAACCGGTTCCGTC)·d(GACGGAACCGGTTCCGTG) and d(GGCGCGCACCACGCGCGG)·d(CCGCGCGTGGTGCGCGCC).

The covariance matrix defining the deformability of helical parameters of a given DNA segment (for example a dinucleotide step) is computed from the ensemble of molecular dynamics simulations and inverted to determine 6 × 6 stiffness matrix for each fragment (for example each of the ten unique dinucleotide steps, or the ten dinucleotide steps adapted to all tetramer environments, which were derived from the ABC consortium):

$$\Theta = k_B T C^{-1} = \begin{pmatrix} k_w & k_{wr} & k_{wt} & k_{ws} & k_{wl} & k_{wf} \\ k_{wr} & k_r & k_{rt} & k_{rs} & k_{rl} & k_{rf} \\ k_{wt} & k_{rt} & k_t & k_{st} & k_{tl} & k_{tf} \\ k_{ws} & k_{rs} & k_{st} & k_s & k_{ls} & k_{lf} \\ k_{wl} & k_{rl} & k_{tl} & k_{ls} & k_l & k_{lf} \\ k_{wf} & k_{rf} & k_{tf} & k_{lf} & k_{lf} & k_f \end{pmatrix}$$

where k_b is the Boltzman constant, T is the absolute temperature, and k stands for the different stiffness constants defining the 36 elements of the stiffness matrix ( ) (twist (w), roll (r), tilt (t), rise (s), slide (l) and shift (f)) at the dinucleotide level obtained by inversion of the MD-associated covariance matrix (C). Pure stiffness constant associated to individual helical deformations (ktilt, kroll, kshift, ktilt, krise and kslide) are taken from the diagonal of the matrix. Ktotal is obtained as the product of the six pure stiffness constants and gives a rough global estimate of the flexibility of each base pair step.

2.5.8 Calculation of Nucleosome Deformation Energy

Stiffness matrix described above was also used to determine *ab initio* (i.e. without any knowledge-based training) the energy required to wrap a 147 bp long DNA sequence into a nucleosome conformation, assuming that distortion is naturally harmonic. This was determined as:

$$E = \frac{1}{2}\Theta(X - X_0)^2$$

where *X* stands for the (helical) geometry of the DNA in the crystal structure of nucleosome, and X_0 stands for the equilibrium geometry of the same sequence of DNA in water in the absence of histones (also obtained from MD). The reference nucleosome structure was obtained by averaging and smoothing of all available X-ray structures of the nucleosome core particle using a Fourier Transform algorithm (Harp et al. 2000; Luger et al. 1997; Ong et al. 2007; Muthurajan et al. 2004; R K Suto et al. 2000; Lavery et al. 2009; Bao et al. 2006). This procedure reduces local variability that can be due to crystallization artifacts. Note that large E energy values signal those regions where physical descriptors indicate that wrapping a DNA in a left-handed superhelix is expected to be difficult, i.e. very likely regions where physical properties of DNA do not favor nucleosome formation.

2.5.9 Statistical Positioning Model

The very simple statistical positioning model featured considers that, after the energetic barrier in the NFR, nucleosomes are arranged statistically with a lineal increasing fuzziness. We decided to simulate a population of nucleosome reads centered at the +1 nucleosome with a dyad deviation of 25bp. Dyads of downstream nucleosomes (+2, +3, …) were spaced 147 + 14bp (accounting for average linker DNA length) with an increasing deviation of the dyad of +5bp in every step and a decreasing number of reads equal to the 4% of the previous peak. The dyad of the -1 nucleosome was placed 147+100bp (247bp in total) upstream the +1 for the closed NFR and 147+200bp (347bp in total) for the open NFR. The following upstream nucleosomes (-2, -3, …) were defined as in the case of the downstream model but adjusting the deviation in 35bp in the -1 nucleosome plus 5bp in every following step, with a linker length of 18bp. Different values of the different parameters in the model were selected after a grid search

maximizing the correlation of the model with the average experimental distribution.

2.5.10 TFBS Prediction

Transcription factor binding sites (TFBSs) were derived from the position weight matrices (PWM) available in JASPAR database for yeast (Bryne et al. 2008). For every PWM, the genome-wide binding scores and predicted TFBS were calculated using R/Bioconductor Biostrings library with default parameters. Regions with annotated TFBS were pooled and their coverage was calculated as a measure of global TF affinity genome-wide.

RESULTS

1 Nucleosome Positioning by MNase-Seq and MNase-CHIP in *S. cerevisiae*

To show the reproducibility of nucleosome positions on a genomic scale and compare different technologies, we generated genome-wide nucleosome maps of *S. cerevisiae* strain BY4741 using both DNA tiling microarray and high-throughput sequencing that identify nucleosomal and linker DNA sequences based on MNase susceptibility. Nucleosomal DNA samples were prepared in the same manner for both platforms, as described in Materials&Methods, 2.1.3. The samples mostly yielding to mononucleosome fragments **(Figure 15)** were labeled as n*ucleosomal DNA*. Naked DNA digested by MNase was used as a control (Materials&Methods, 2.1.1& 2.1.2) and labeled as *genomic DNA*.

For the tiling array approach (MNase-CHIP), nucleosomal and genomic DNA samples were labeled fluorescently with Cyanine (Cy) 3 and Cy5, respectively by IRB Functional Genomics Core Facility. The labeled samples were then hybridized to *S. cerevisiae* Tilling 1.0R Array (Affymetrix), which includes 5,744 probes of 25-mers tiled every 5 bp across the *S. cerevisiae* genome (Materials&Methods, 2.5.1). Nucleosome position signals were detected based on the hybridization log ratio between nucleosomal and genomic DNA, which were converted to coverage maps of 1 bp resolution by the nucleR algorithm. Here, the coverage represents the normalized fluorescence intensity from hybridized nucleosomal versus genomic DNA samples. Particularly, we identified nucleosome occupied regions as those with a ratio higher than the average while nucleosome free regions displayed a ratio lower than the average.

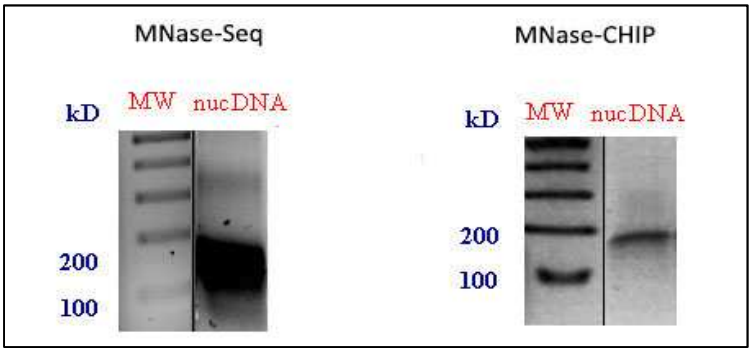


Figure 15: Digestion levels of nucleosomal DNA. The isolated DNA from MNase-digested chromatin was run on 2% agarose gel. On the left (right) side, the nucleosomal DNA from MNase-Seq (MNase-CHIP) experiment is shown.

On the other hand, for high-throughput sequencing approach (MNase-Seq), nucleosomal and genomic DNA samples were single-end sequenced, on an Illumina/Solexa Genome Analyzer (GA) IIx. Data processing with GA base calling pipeline yielded 7.6 million reads of 38 bp length. Once the reads were obtained, they were aligned to the reference yeast genome (SaccCer3 (2011)). The coverage was calculated for each base pair, representing the read counts at each position. Finally, to obtain the nucleosome peaks, the rea

ds were trimmed and extended to 50bp around their center, providing a fixed read-width of 50bp in order to remove the potential noise. The schematic pipeline of data processing and peak detection is shown in **Figure 16A**. All the peaks above a nucleR score of 0.2 are considered as nucleosome positions. We can find a nucleosome read virtually in any place in the genome, however some positions will show higher concentration within a cell population and we define these nucleosomes as “well-positioned”; while some others will be less phased and may occupy different locations on the genome within a population and those are defined as “fuzzy” nucleosomes. We determined the phasing of the nucleosomes according to their nucleR score, which is based on height (i.e. direct measurement of read coverage) and sharpness (measurement of fuzziness) of the peak. Accordingly, a very narrow peak that is not surrounded by other peaks represents a well-positioned nucleosome, whereas wide peaks or peaks very close to each other are fuzzy nucleosomes (**Figure 16B**).

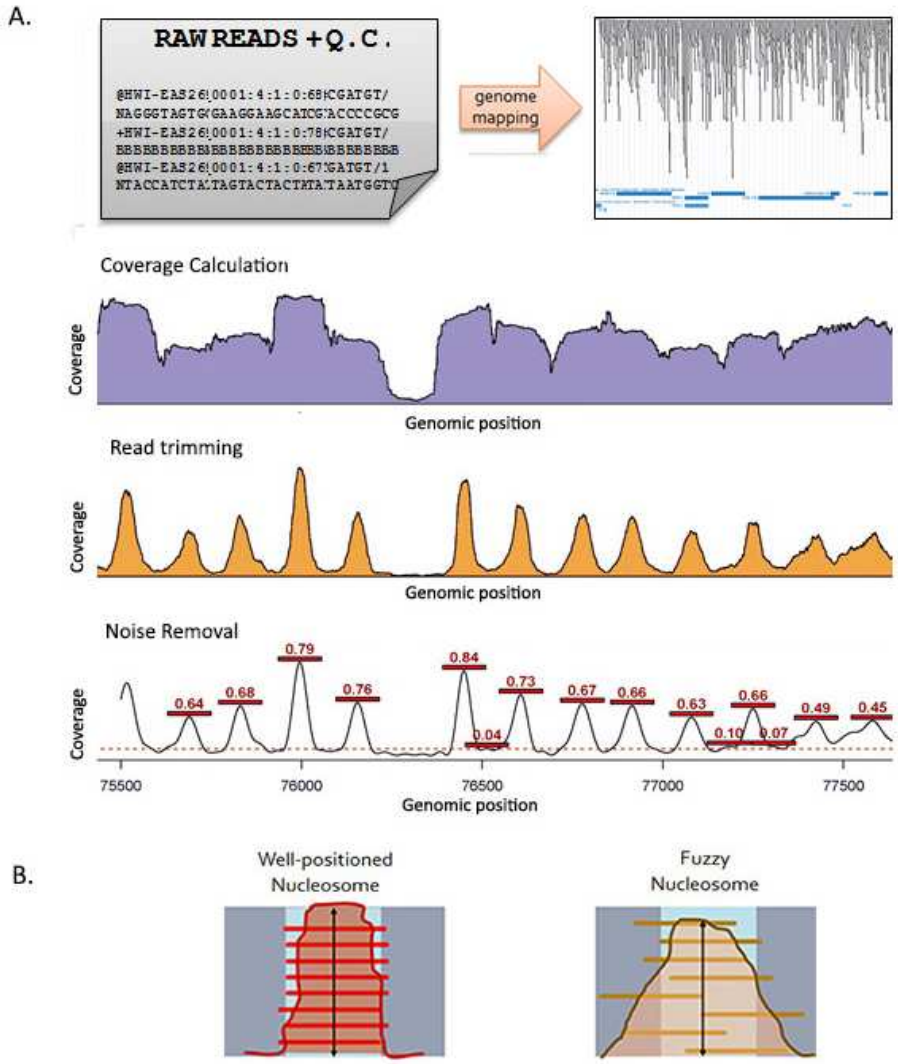


Figure 16: Data analysis and peak detection of nucleosome positioning experiments. Peak detection and nucleosome calling were performed in a similar way for both tiling microarray and sequencing experiments.

Consequently, we detected in total 77,360 peaks, of which 54,164 are annotated as well-positioned and 23,196 are fuzzy nucleosomes in tiling microarray experiment, whereas 51,857 well-positioned and 58,946 fuzzy nucleosomes were detected in high-throughput sequencing experiments (**Table 1**).

Platform	Resolution	Coverage	Peak Count	W Nucl.	F Nucl.
Illumina Solexa 2G	1 bp	24x *1,076,056 uncovered bases*	110,803	51,857	58,946
Affymetrix 2.0 (5 bp)	5 bp	*83,799 uncovered base*	77,360	54,164	23,196

Table 1: Comparison of nucleosome maps generated using two distinct platforms. The platforms are compared regarding the resolution, coverage obtained along the genome, number of peaks detected and number of well-positioned and fuzzy nucleosomes annotated.

When compared to MNase-CHIP, MNase-Seq is able to detect more nucleosomes along the genome (**Table 1**). This might be explained by the higher resolution of the sequencing platform or by the probe dependency of arrays. Since the regions covered by MNase-CHIP depend on the probe hybridization, the 5,744 probe set provides information for around 5,750 genes. On the other hand, MNase-Seq covers the whole genome in a single base pair resolution. Finally, as repetitive regions cannot be removed from the tiling array data, the uncovered bases are much fewer in MNase-CHIP maps, although MNase-Seq gives more coverage and precise information about the most relevant non-repetitive regions.

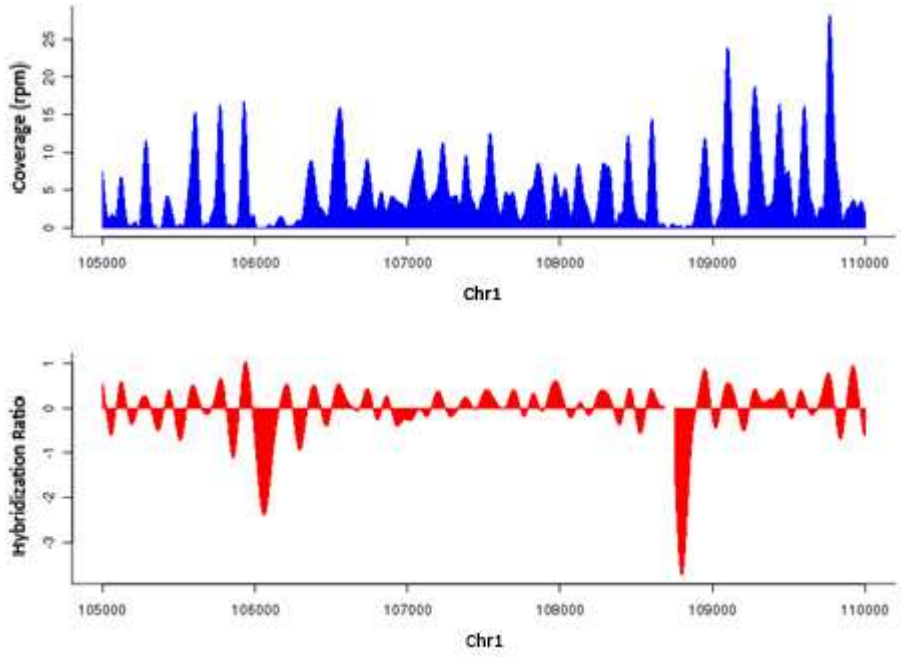


Figure 17: Coverage profile of MNase-Seq and MNase-CHIP nucleosome maps. 105 000- 110 000 bp region of chromosome I is demonstrated for the comparison of MNase-Seq (top) and MNase-CHIP (bottom) nucleosome profiles.

1.1 Nucleosome Positioning around TSS and TTS

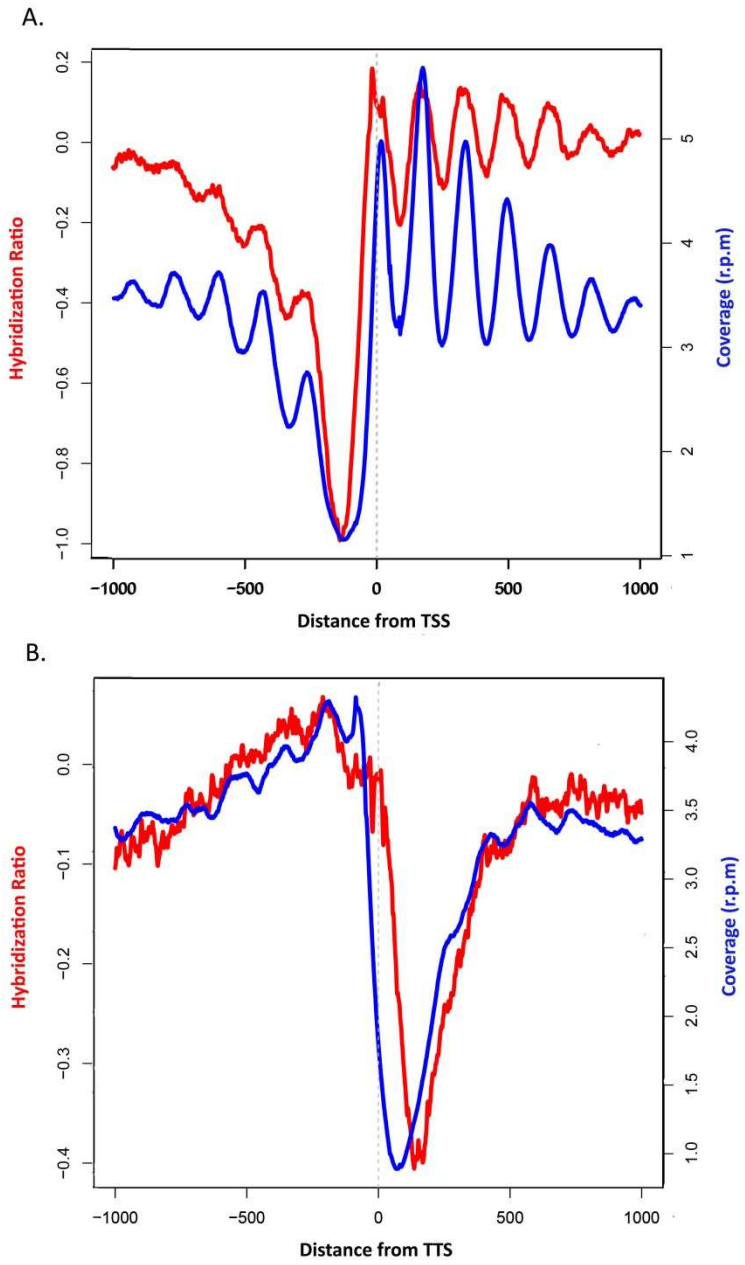


Figure 18: Nucleosome organization around TSS (A.) and TTS (B.) in two datasets. Blue (red) line represents the nucleosome distribution in the nucleosome map generated by MNase-Seq (MNase-CHIP).

When nucleosome maps of two platforms are compared, both profiles are consistent and the majority of nucleosome depleted and occupied regions are identical (**Figure 17**). As reported earlier, both datasets display nucleosome depletion at intergenic regions as compared with coding regions. In an attempt to compare the promoter regions, around 5,750 promoters were aligned by the start codons to map the coverage around TSSs. Similarities among the maps are especially notable at TSS regions, where both nucleosome profiles exhibit the expected canonical architecture, consisting of 5' end NFR flanked by well-positioned nucleosomes (**Figure 18**). Even though the general profile is quite similar, there are some inconsistencies among MNase-CHIP and MNase-Seq nucleosome maps. While MNase-Seq maps have sharper nucleosome peaks, in MNase-CHIP the peaks are wider and less defined, especially the nucleosome signals upstream TSSs, which are poorly separated from each other. This might be due to the distinct resolution of the platforms. Furthermore, NFR width is around 50 bp narrower in MNase-CHIP than in MNase-Seq maps, which is similar to the width of NFRs in previous studies (Lee et al. 2004; Mavrich et al. 2008). Finally, another notable difference among nucleosome maps is the downstream nucleosome positioning. In MNase-Seq maps, nucleosomes over the coding regions are arranged in a regularly spaced array, starting at +1 nucleosome position and decrease in phasing with the distance from TSS, as reported earlier (Albert et al. 2007; Mavrich et al. 2008; Sadeh & Allis 2011). On the other hand, such decay in phasing is not observed in downstream nucleosomes of MNase-CHIP maps, which have strong positioning instead, as demonstrated in other nucleosome profiles generated by tiling microarray (Lee et al. 2007).

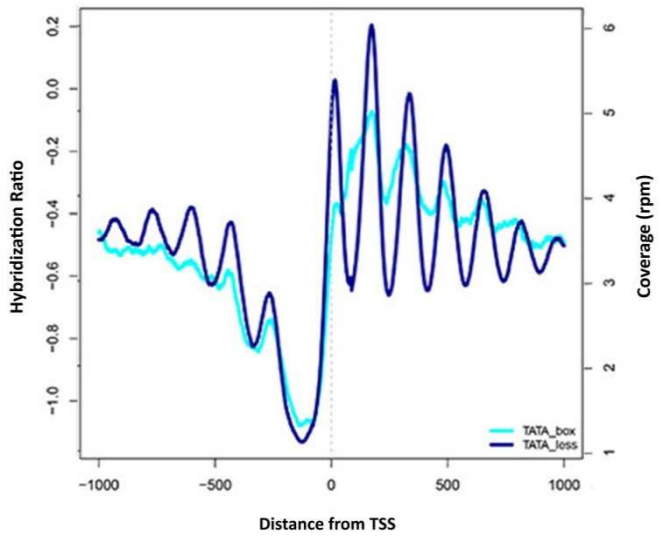
Like TSSs, nucleosome profiles around transcription termination sites (TTSs) show high similarities in the two datasets, even though the MNase-CHIP map contains a higher noise level. In both maps, the downstream TTS is depleted of nucleosomes, which is flanked by high nucleosome position signals. However, unlike TSS nucleosome signals, TTS signals are poorly defined in both maps. Remarkably, there is a slight shift towards downstream in MNase-CHIP signals. Moreover, 3' NFR width is distinct in two datasets;

MNase-Seq dataset NFR is around 120 bp wide while MNase-CHIP dataset has shorter NFR, around 100 bp (**Figure 18**).

1.2 Nucleosome Positioning at Different Promoter Types

In an attempt to compare nucleosome profile of different promoter types, we examined the nucleosome profile of TATA-less and TATA-box containing genes, annotations uploaded from Basehoar et al. (Basehoar et al. 2004).

A.



B.

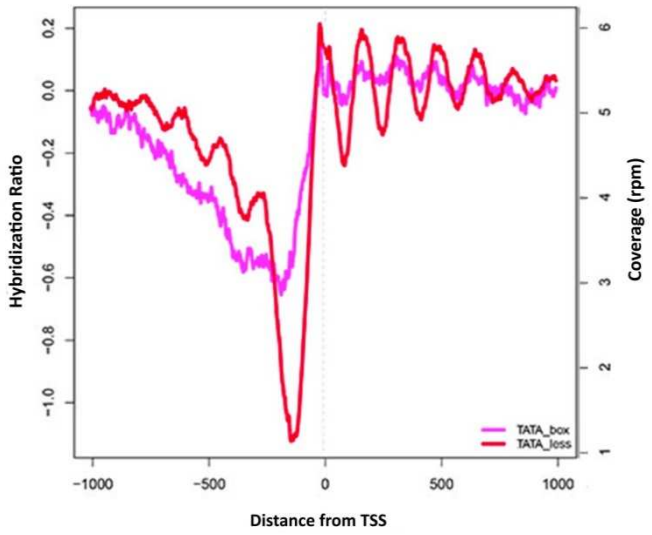


Figure 19: Nucleosome organization around TSSs of TATA-less and TATA-box containing genes for A) MNase-Seq, B) MNase-CHIP nucleosome maps. TATA-box annotations are uploaded from (Basehoar et al. 2004).

Nucleosome profiles of 1073 TATA-box genes and 4471 TATA-less genes areplotted for both CHIP and Seq datasets (**Figure 19**). Consistent with the previous reports (Zaugg & Luscombe 2012), TATA-box genes follow distinct nucleosome profiles compared to TATA-less genes in both maps; NFR is shallower and narrower, very similar to the pattern of closed promoters. Moreover, TATA-box genes exhibit lower coverage and less nucleosome phasing, especially upstream TSS. The phasing of nucleosomes is already lost after +4 position or before -1 nucleosome. As the nucleosome signals are more diffused in TATA-less genes, the linker regions are less defined as well in both data sets.

Taken together, even though the general nucleosome pattern is very similar in both MNase-CHIP and MNase-Seq nucleosome maps, we performed the rest of our analysis based on the nucleosome maps using MNase-Seq approach, due to higher resolution, lack of probe dependency and the possibility of obtaining precise information about non-repetitive region.

2 Impact of DNA Physical Properties on Nucleosome Positioning

DNA sequence has been considered to be an important contributor to nucleosome assembly (Chung & Vingron 2009; Kaplan et al. 2009; Ioshikhes et al. 2006). However, crystal structures of nucleosome core particles reveal a lack of direct read-out mechanisms between histone proteins and DNA bases (Luger et al. 1997; Richmond & Davey 2003). This led to postulate that the DNA sequence relative affinities for nucleosome formation (e.g. high-affinity Widom601 sequence) (Lowary & Widom 1998) might be actually based on an indirect read-out mechanism to modify equilibrium geometry of a given DNA sequence to adopt the nucleosome-bound geometry. Moreover, indirect evidence has highlighted the interplay between DNA physical properties and chromatin organization (Goñi et al. 2008; Miele et al. 2008). To pursue the role of DNA physical properties on nucleosome positioning, we have applied a genome-wide approach comparing MNase digestion profiles of yeast genomic DNA and nucleosomal DNA.

2.1 Preferential MNase Cut Sites

We prepared genomic and nucleosomal DNA samples as described in Materials&Methods, 2.1. In order to obtain a comparable fragmentation level with nucleosomal DNA, genomic DNA was partially digested with MNase to a range of 100-400 bp (**Figure 20**). To question whether MNase has sequence-specific cleavage, we first compared MNase cut sites of genomic and nucleosomal DNA based on the resulting sequencing data. MNase cut sites were defined in tetramers by taking two bases upstream and downstream of each read end (**Figure 21**).

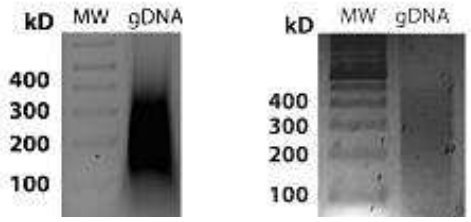


Figure 20: Digestion levels of nucleosomal DNA. The isolated DNA from MNase-digested chromatin (left) and sonicated chromatin (right) was run on 2% agarose gel.

For the comparison, we calculated the ratio between the experimentally detected and the expected tetramer frequencies, which were calculated by sampling ten million tetramers in the entire yeast genome. This ratio was then used to identify the over- or under-represented MNase cut sites.

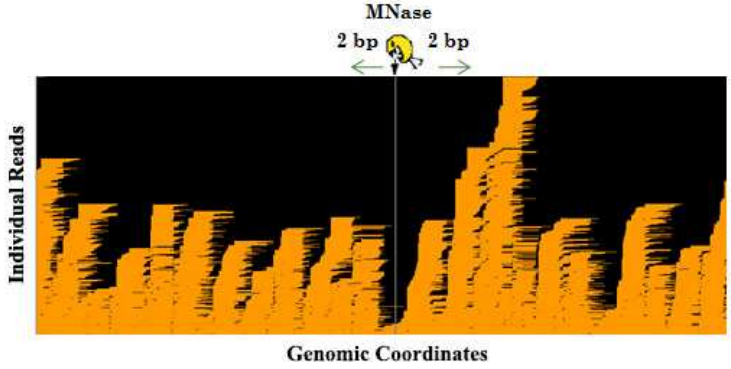


Figure 21: Definition of MNase cut sites. MNase cut sites were defined in tetramers by taking two bases upstream and downstream of each read end.

The over-represented MNase cut sites are shown in **Table 2**. We observe that in both genomic and nucleosomal DNA, the MNase preferentially cuts tetramers with a central d(A-T) step, but without the requirement of flanking

dC or dG bases, in contrast to previous low-scale experiments (Flick et al. 1986). The high-cutting susceptibility for d(CATA)·d(TATG) tetramers found in mouse satellite DNA (Hörz & Altenburger 1981) is also detected in the genomic DNA sample, although these tetramers are not the most predominant cutting sites. Overall, MNase displays quite strong sequence preferences in both samples, suggesting that intrinsic susceptibility to MNase of genomic DNA can bias the cutting scheme in nucleosomal DNA.

Genomic DNA	**ratio**	**p-val**	**Nucleosomal DNA**	**ratio**	**p-val**
TATA.TATA	13.28	$< 10^{-18}$	CTAG.CTAG	4.07	$< 10^{-18}$
ATAG.CTAT	8.45	$< 10^{-18}$	ATAG.CTAT	3.93	$< 10^{-18}$
CTAA.TTAG	7.90	$< 10^{-18}$	CAAG.CTTG	3.57	$< 10^{-18}$
CTAG.CTAG	6.80	$< 10^{-18}$	CTTA.TAAG	3.52	$< 10^{-18}$
ATTA.TAAT	5.74	$< 10^{-18}$	CATG.CATG	3.42	3.01×10^{-4}
CATA.TATG	5.62	$< 10^{-18}$	CATA.TATG	3.11	$< 10^{-18}$
ATAA.TTAT	5.14	$< 10^{-18}$	CTAA.TTAG	3.00	$< 10^{-18}$
CTTA.TAAG	4.92	$< 10^{-18}$	CTAC.GTAG	2.98	$< 10^{-18}$
TTAA.TTAA	4.64	$< 10^{-18}$	ATTG.CAAT	2.96	$< 10^{-18}$
ATAT.ATAT	4.52	$< 10^{-18}$	AAAG.CTTT	2.82	$< 10^{-18}$
TAAA.TTTA	3.48	$< 10^{-18}$	CTTC.GAAG	2.79	$< 10^{-18}$
ATTG.CAAT	3.25	$< 10^{-18}$	AATG.CATT	2.50	$< 10^{-18}$
GTAA.TTAC	2.64	1.01×10^{-4}	CATC.GATG	2.24	6.03×10^{-4}
ATAC.GTAT	2.39	2.01×10^{-4}	CAAC.GTTG	2.19	10^{-3}
			CAAA.TTTG	2.17	$< 10^{-18}$

Table 2: MNase-preferred tetramer frequencies at the cutting sites. Experimentally detected and expected frequency ratios of MNase-preferred tetramers for genomic (left) and

nucleosomal (right) DNA are shown. The significance (p-value) of the enrichment or depletion was calculated for ten million random observations.

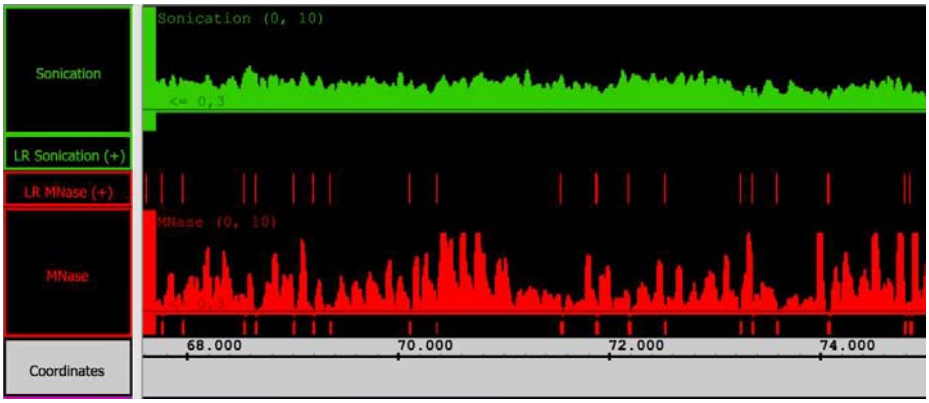


Figure 22: Coverage profiles sonicated genomic DNA and MNase-digested genomic DNA. The sonicated genomic DNA (green) and MNase-digested DNA (red) profiles are shown for chromosome 16 in region between 68 000 – 74 000 bp.

To see whether the high sequence preferences could be an experimental artifact or sequencing bias, we performed a control experiment by fragmenting the same genomic DNA by sonication to a range of 100-400 bp (**Figure 20**). As shown in **Figure 22**, the sonicated genomic DNA (green profile) does not show any marked variation in the coverage profile, whereas the MNase-digested sample (red profile) clearly shows more favored cleavage regions. Moreover, among 14 tetramer sequences with high frequencies in genomic DNA, 6 are common with nucleosomal DNA, showing a good agreement in the preferred cutting sites between naked and nucleosomal DNAs. This suggests that preferred tetramer signals that are directing the first MNase cut in chromatin are not random and are intrinsic to the genomic DNA sequence.

On the other hand, tetramers resistant to MNase cleavage are very diverse, except for the presence of a central purine-purine dinucleotide step (**Table 3**). Moreover, they are different between naked and nucleosomal DNA samples, demonstrating that the nucleosome structure might specifically protect certain sequences from MNase degradation, which are not observed in genomic DNA.

Genomic DNA	ratio	p-val	Nucleosomal DNA	ratio	p-val
AACT.AGTT	0.062	10^{-4}	AGGA.TCCT	0.062	$4.00x10^{-4}$
GGAA.TTCC	0.078	$< 10^{-18}$	AGCA.TGCT	0.062	$2.00x10^{-4}$
AGAT.ATCT	0.078	$< 10^{-18}$	ACCT.AGGT	0.065	$1.51x10^{-3}$
ACCA.TGGT	0.082	$< 10^{-18}$	AAGC.GCTT	0.097	$6.00x10^{-4}$
AAGT.ACTT	0.103	$< 10^{-18}$	TCCA.TGGA	0.098	10^{-4}
AGAA.TTCT	0.109	$< 10^{-18}$	ACCA.TGGT	0.121	$9.00x10^{-4}$
AAGA.TTCT	0.117	$< 10^{-18}$	AAGG.CCTT	0.196	$1.11x10^{-3}$
ATCA.TGAT	0.149	$< 10^{-18}$	AAGA.TCTT	0.211	$< 10^{-18}$
TGAA.TTCA	0.229	$< 10^{-18}$	AAGT.ACTT	0.227	$1.21x10^{-3}$
AAAA.TTTT	0.251	$< 10^{-18}$	AACA.TGTT	0.249	$1.21x10^{-3}$

Table 3: MNase non-preferred tetramer frequencies at the cutting sites. Experimentally detected and expected frequency MNase-non-preferred tetramer ratio for genomic (left) and nucleosomal (right) DNA are shown.

2.2 Preferential MNase Degraded Regions

Upon an initial endonucleotic cleavage, MNase displays an exonuclease activity that continues with the digestion of DNA (Alexander et al. 1961). We defined those extensive MNase digested sites as low coverage regions (LRs), which are determined in genomic DNA as regions within the bottom 2.5 percentile of the sample coverage, excluding non-zero coverage regions, and in nucleosomal DNA as regions within the bottom 10 percentile regarding majority of depleted regions intrinsically caused by the nucleosome free regions. Explicitly, LRs of 5 and 250 bp in length were selected for analysis, and the ones in a distance shorter than 4 bp were merged. The LRs identification is schematically shown in **Figure 23**.

The analysis of LRs reveal that LR tetramers of genomic and nucleosomal DNA are mainly composed of d(A-T)s, which have weaker hydrogen bonds

(**Table 4**). Even though some tetramers are common in MNase cut sites and LRs (**Table 2, Table 3**), some tetramers are unique to LRs, suggesting that the digestion of a particular fragment does not only depend on the vicinity of the cleavage site, but also on the differential sequence preferences of endo- and exo-nuclease activities. For example, while d(AAAA·TTTT) is nearly four times more frequent than expected ($p< 10^{-8}$), it is rarely present at primary cutting sites (1/4 times less than expected, $p< 10^{-7}$;**Table 2, Table 4**). Moreover, not only tetramer composition but also LR

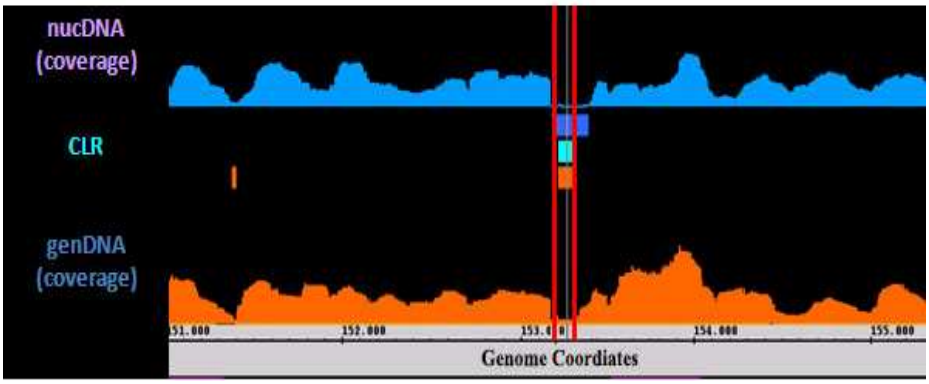


Figure 23: LRs and CLRs determination. Low regions (LRs) are determined in genomic and nucleosomal DNA as regions within bottom 2.5 and 10 percentile, respectively. CLRs are determined as intersection of genomic and nucleosomal DNA LRs and shown as light blue box.

locations are very similar in genomic and nucleosomal DNAs. Therefore, we have defined the intersection of LRs as common low regions (CLRs) for further analysis (**Figure 23**). 2,770 regions were identified, which comprise 57.60% of genomic DNA LRs. This high intersection indicates that sequence susceptibility of MNase digestion in nucleosomal DNA is not exclusively dependent on the chromatin structure, but also related to the DNA intrinsic properties (**Table 4**).

Genomic DNA	ratio	p-val	Nucleosomal DNA	ratio	p-val	Common low regions (CLR)	ratio	p-val
AAAA.TTTT	3.87	$< 10^{-18}$	TATA.TATA	4.06	$< 10^{-18}$	AAAA.TTTT	4.48	$< 10^{-18}$
TAAA.TTTA	2.38	$< 10^{-18}$	ATAT.ATAT	3.09	$< 10^{-18}$	TATA.TATA	3.18	$< 10^{-18}$
TATA.TATA	2.38	9.1×10^{-4}	AAAA.TTTT	2.91	$< 10^{-18}$	TAAA.TTTA	2.67	$< 10^{-18}$
AAAT.ATTT	2.16	$< 10^{-18}$	ATAA.TTAT	2.21	$< 10^{-18}$	ATAA.TTAT	2.62	$< 10^{-18}$
ATAA.TTAT	2.13	$< 10^{-18}$	AATA.TATT	2.08	$< 10^{-18}$	ATAT.ATAT	2.57	$< 10^{-18}$
TTAA.TTAA	2.10	7.5×10^{-3}	ATTA.TAAT	1.99	10^{-4}	AATA.TATT	2.43	$< 10^{-18}$
AATA.TATT	2.02	$< 10^{-18}$	TAAA.TTTA	1.84	7.0×10^{-4}	TTAA.TTAA	2.29	3.2×10^{-3}
ATAT.ATAT	2.00	4.6×10^{-3}	AAAT.ATTT	1.62	4.2×10^{-3}	AAAT.ATTT	2.27	$< 10^{-18}$
AATT.AATT	1.84	5.5×10^{-3}				ATTA.TAAT	2.15	$< 10^{-18}$
ATTA.TAAT	1.79	3.6×10^{-3}				AATT.AATT	1.81	1.3×10^{-2}
GAAA.TTTC	1.45	3.4×10^{-2}						

Table 4: Tetramer frequency of MNase-digested LRs and CLRs. Experimentally detected and expected tetramer frequency ratios in genomic (left) and nucleosomal (center) DNA LRs, and in CLRs (right) are shown.

2.3 Low Coverage Regions and Physical Properties

The diversity among MNase tetramer cut sites and digested regions suggest a possible contribution of the indirect read-out mechanism in MNase digestion preferences on chromatin. To explore this possibility, we analyzed Molecular Dynamics-derived physical properties, calculated as described in Materials&Methods, 2.5.7. We compared these predicted properties at preferential MNase cut sites versus non-preferential ones in genomic DNA.

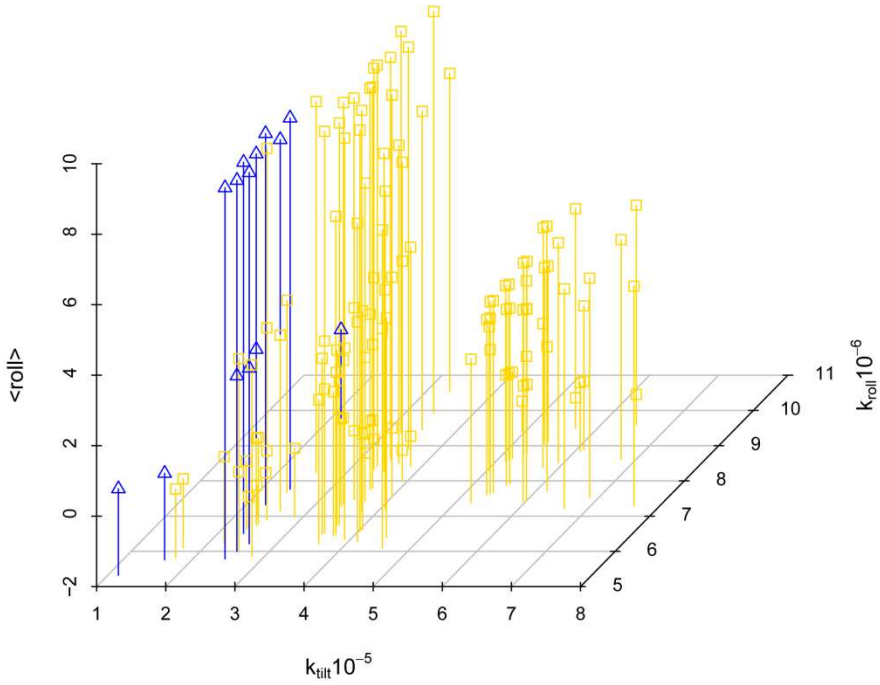


Figure 24: Physical properties at MNase preferred and non-preferred cut sites. Representation of preferential (blue-triangles) vs. non-preferential (yellow-squares) MNase cutting sites in genomic DNA with respect to physical properties tilt and roll stiffness (in kcalmol-1degree-2), and equilibrium roll (in degrees) for each tetramer are shown.

Preferred MNase cut sites (blue arrows; **Figure 24**) have lower k_{tilt} and k_{roll} values, which are characteristics of high flexibility. Additionally, higher roll values are strikingly observed in these sites, indicating a wide opening in the major groove at the equilibrium geometry. On the other hand, such pattern is not observed in the non-preferential sites, suggesting that certain sites are more accessible to MNase cleavage, determined by their physical properties (**Figure 24**).

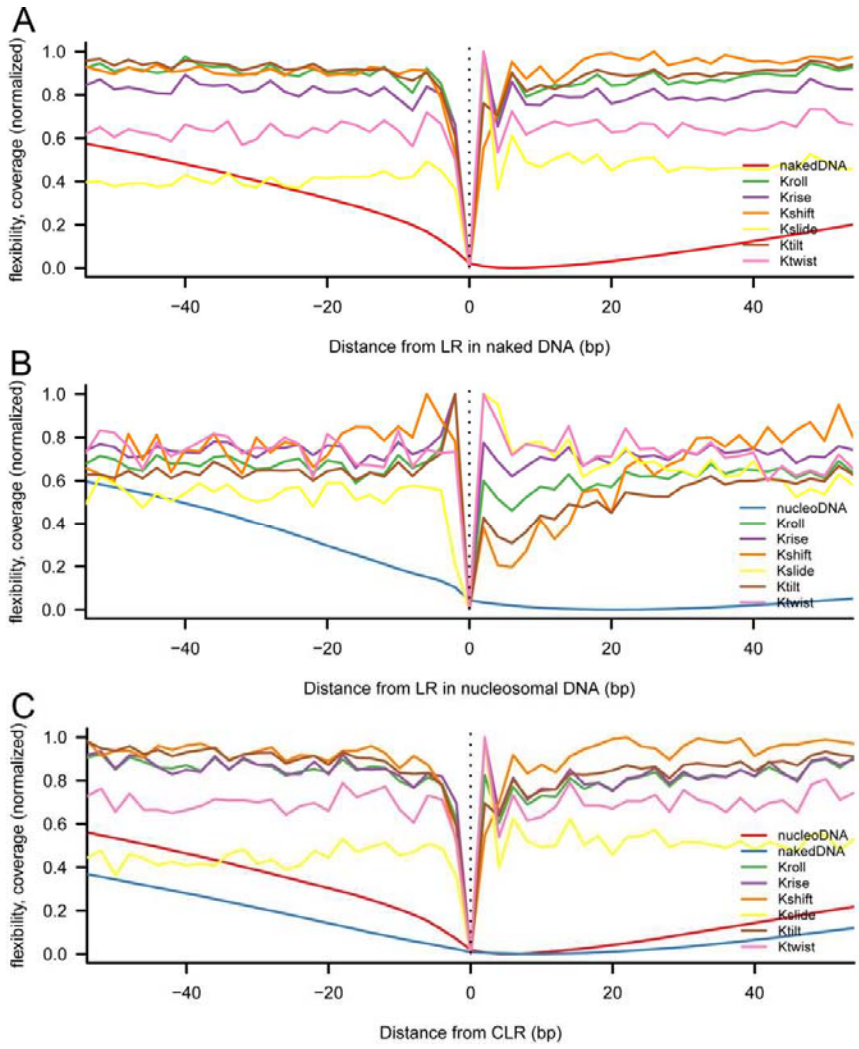


Figure 25: Individual Stiffness profiles in low coverage regions. Six individual stiffness parameters (k_{roll}, k_{tilt}, k_{twist}, k_{shift}, k_{rise} and k_{slide}) and coverage maps were calculated and averaged across all yeast genome, around genomic (A), nucleosomal (B)DNA LRs and CLRs (C). All values are normalized (in the 0-1 range) to facilitate analysis and comparisons.

As a next step, we plotted each dinucleotide-based stiffness parameter around genomic and nucleosomal DNA LRs and CLRs to investigate possible impact of physical properties on MNase exonuclease digestion. Interestingly, there is a sudden decrease in all stiffness parameters at LRs and CLRs, while the flanking regions have higher values. As stiffness parameters anti-correlate with DNA flexibility, this observation at LRs and CLRs suggests these regions to be highly flexible and surrounded by stiff motifs (**Figure 25**).

2.4 Nucleosome Positioning and Gene Structure

In an attempt to explore the positioning of MNase sensitive and resistant regions, we plotted the coverage profiles of genomic and nucleosomal DNA around TSSs and TTSs (**Figure 26**). As reported previously (Lee et al. 2007; Mavrich et al. 2008; Kaplan et al. 2009; Ioshikhes et al. 2006), MNase resistant regions in nucleosomal DNA are mainly concentrated at the beginning of coding regions and upstream TTSs, as indicated as blue dash lines in **Figure 26A&B**. The same regions are found to be MNase resistant in genomic DNA, indicated as red lines. On the other hand, MNase sensitive regions in nucleosomal and genomic DNA are mostly detected at upstream TSSs or downstream of TTSs, which are defined as 5' and 3' NFR in nucleosome maps. Remarkably, such variations in the genomic DNA coverage profiles are not observed in the sonication-fragmented sample (green lines), indicating that the MNase observed variations are not due to experimental or processing artifacts, as mentioned before.

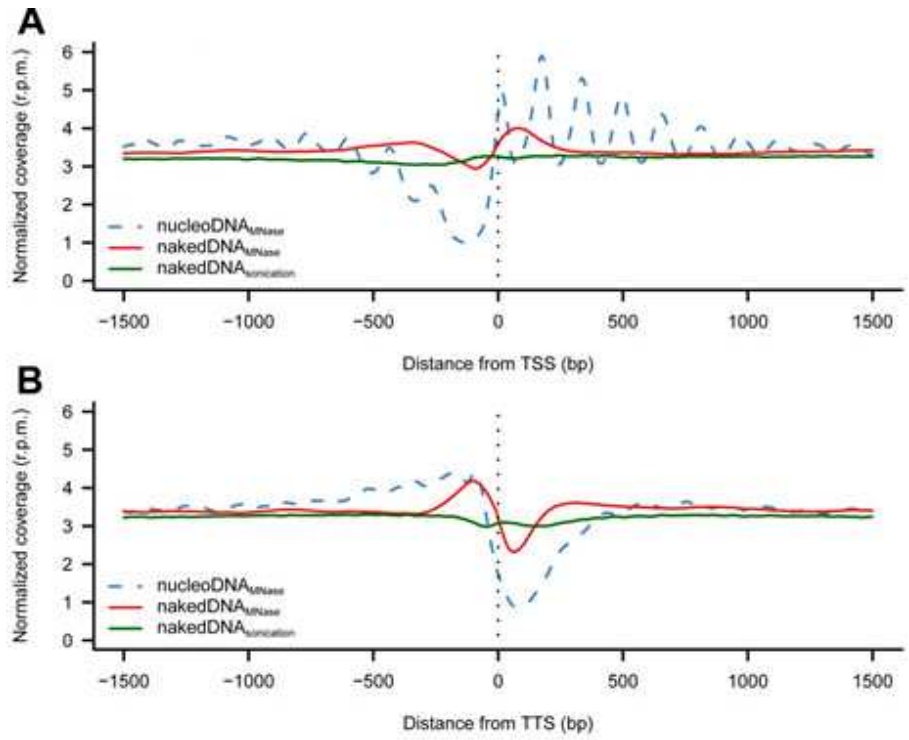


Figure 26: Coverage profiles of genomic and nucleosomal DNA. Coverage maps per base pair were calculated and averaged across all yeast genome around TSSs (A) and TTSs (B) for MNase-digested genomic and nucleosomal and sonicated genomic DNAs.

The similarities observed between nucleosomal and genomic DNA profiles indicate that nucleosome maps might not only reflect nucleosome positioning, but also the intrinsic susceptibility of genomic DNA to MNase digestion. The analogy between nucleosomal and genomic DNA profiles is clearly illustrated in the corrected nucleosomal DNA profile, in which there is a reduction in the nucleosomal signal around TSSs and TTSs (**Figure 27**). Despite the decrease in the signal, the general profile does not change, pointing out that MNase digestion does not lead to any significant bias in nucleosomal maps, but it might digest similar regulatory regions in the genomic and nucleosomal DNA due to the intrinsic susceptibility of DNA.

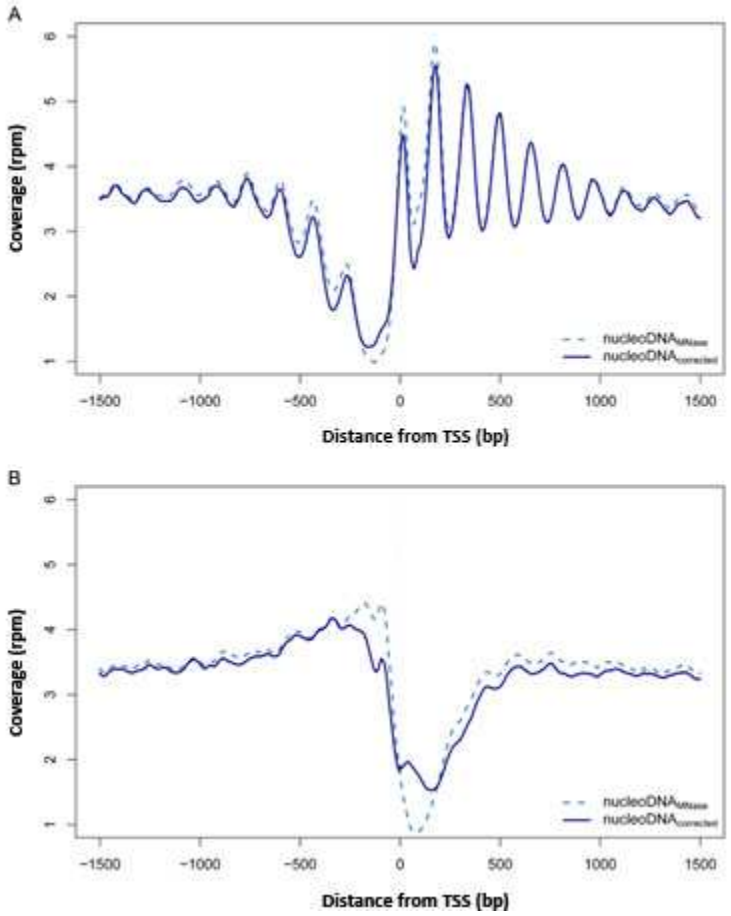


Figure 27: Corrected TSS and TTS coverage profiles. Coverage profiles at transcription start sites (TSSs) (top) and transcription termination sites (TTSs) (bottom) in MNase-digested nucleosomal DNA before (dashed lines) and after naked DNA correction (continuous lines).

2.5 Physical Properties Influence Nucleosome Positioning at TSSs and TTSs

To investigate the impact of DNA intrinsic properties in regulatory regions, the stiffness constants for each parameter are plotted around TSSs and TTSs (**Figure 28)**. Accordingly, upstream TSSs and downstream TTSs are marked by

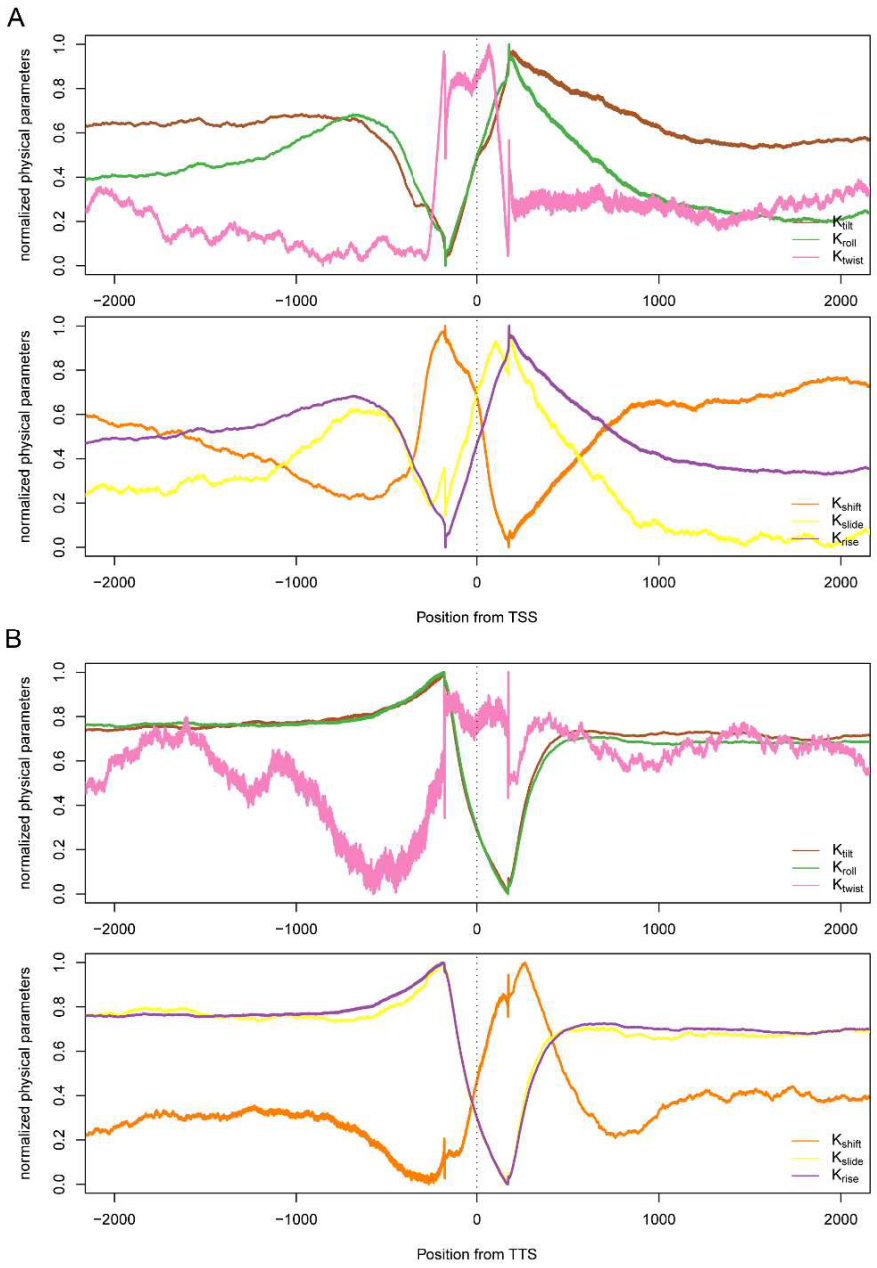


Figure 28: Stiffness parameters at TSSs and TTSs. Plots show the average variation of stiffness parameters (translational or rotational) around TSSs and TTSs in the yeast genome (5,750 genes were considered).

unusual properties, where k_{roll}, k_{tilt}, k_{slide} and k_{rise} are low and k_{twist} and k_{shift} are high. These values indicate that those particular regions are highly flexible and strongly nucleosome depleted, as indicated in **Figure 26.** Therefore, unusual physical properties might control nucleosome positioning in those regions. To verify this hypothesis, we computed the deformation energy required to wrap a DNA sequence around a histone octamer by using a simple elastic energy function based on the MD-derived physical descriptor, as explained in Materials&Methods, 2.5.7 and plotted at CLRs.

CLRs, which are nucleosome depleted and contain a high flexible (4 mer) step, correlate with high deformation energy (**Figure 29**), confirming that these regions are more difficult to wrap around a histone core due to a high-energy cost; hence nucleosome formation is less feasible. Overall, the correlation between CLRs and high deformation energy suggests that the intrinsic properties, e.g. point flexibility that make a DNA segment a good substrate for MNase, are also those that avoid DNA wrapping around a nucleosome.

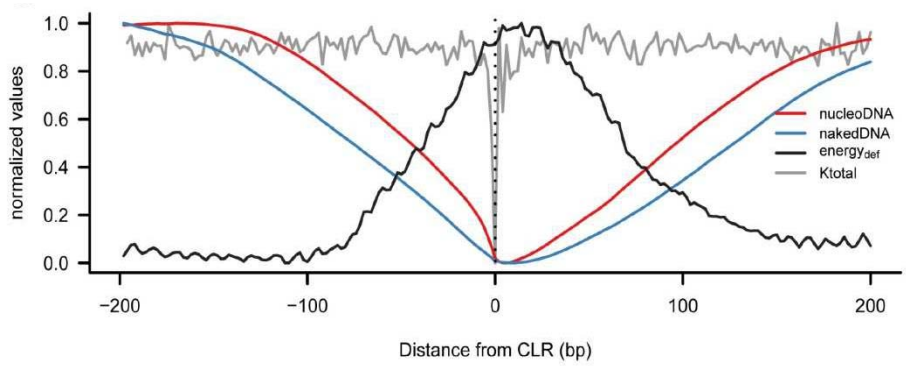


Figure 29: Stiffness, deformation energy and coverage profiles in CLRs. Total stiffness parameter (k_{total}), deformation energy and coverage maps were calculated and averaged across all yeast genome around CLRs. Deformation energy describes the energetic cost of wrapping a 147 bp DNA fragment into the nucleosome conformation.

If the physical properties play a role in determining MNase accessibility and nucleosome formation, deformation energy, which is based on DNA intrinsic properties, should predict *in vivo* nucleosome distribution in yeast. To test this, we plotted the average deformation energy and compared it with

nucleosome profile around TSSs (**Figure 30**). We observe that nucleosome signal is lower where deformation energy is higher and *vice versa*, hence the anti-correlation between deformation energy and coverage profiles illustrates that deformation energy is able to predict nucleosome signals at TSSs. These results suggest that, without dismissing the importance of cellular mechanisms directing chromatin structure, particular features of nucleosome organization around TSSs and TTSs can be rationalized considering physical properties of the genomic DNA sequence.

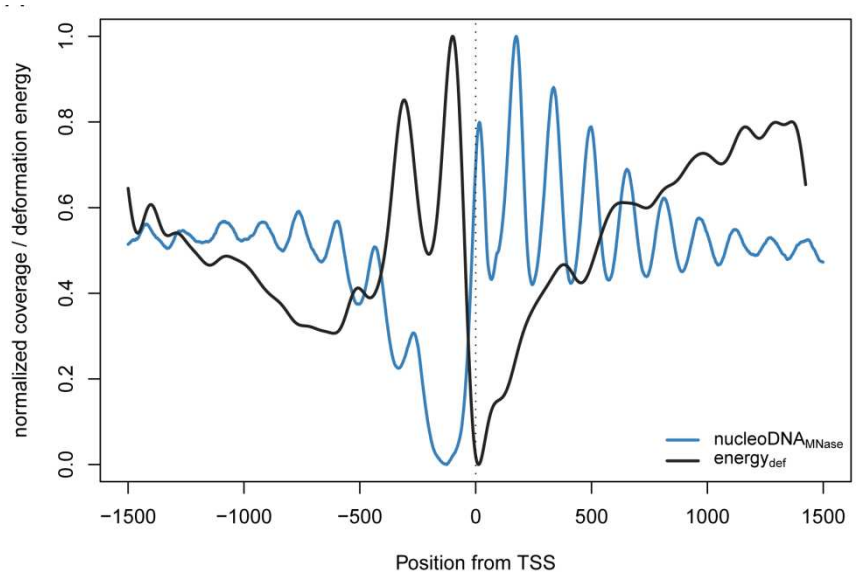


Figure 30: Nucleosome deformation energy around TSSs. Plot displays the average nucleosome deformation energy and nucleosome coverage profile around TSSs. This average is calculated over 5,750 genes.

This study has been published in BMC GENOMICS in 2011 (PMID: 21981733).

3 Fuzziness and Noise in Nucleosome Positioning

To date, genome-wide nucleosome positioning maps have been generated from many model organisms (Kaplan et al. 2008; Lee et al. 2004; Lee et al. 2007; Mavrich et al. 2008; Schones et al. 2008). Each of these maps display quite similar chromatin structure around TSSs. However, individual nucleosome positioning might differ remarkably, so that well-positioned nucleosomes detected in one study might show a fuzzy position or be simply absent in another (Huebert et al. 2012; Bai & Morozov 2010; Kuan et al. 2009; Tsankov et al. 2010). Discrepancy between nucleosome maps may originate from different sources: i) the experimental conditions (such as MNase digestion levels or sequencing protocol); ii) data processing to nucleosome calling; iii) heterogeneity of samples, derived from diversity of cellular states in the culture; and iv) nucleosome dynamics across the genome, that will be detected as positional "fuzziness" in the experimental nucleosomal map (Belch et al. 2010; Lehner 2010). In an attempt to measure the effect of noise in nucleosome occupancy and positioning, we inspected the extrinsic factors that may induce diversity by the comparison of MNase-Seq derived nucleosome maps generated under distinct conditions.

3.1 The Effect of Biological Replica Variability

Firstly, to eliminate the noise resulting from cell population heterogeneity, we have synchronized yeast cultures at the late G1 cell cycle phase. Two synchronized cultures **(Figure 31A**) were considered as biological replicas. We labeled them as *replica* 1 and *replica* 2 and subsequently isolated their nucleosomal DNA under similar MNase digestion conditions. The digestion level was determined by agarose gel electrophoresis and the microfluidics-based platform *Bioanalyzer* (Agilent), as shown in **Figure 31B**. Both samples yielded a major peak around 147 bp corresponding to mononucleosomes, a secondary defined peak around 295 bp corresponding to dinucleosomes and residual peaks at around 60 bp that might be assigned to either tetrasomes or other DNA-protein complexes.

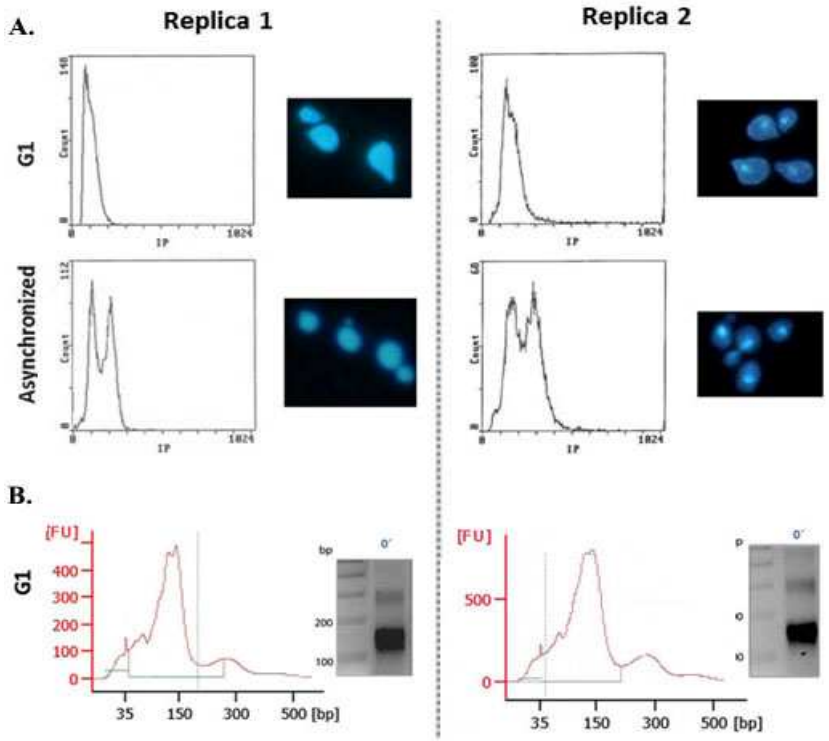


Figure 31: Synchronization and MNase digestion of replica 1 and 2. A. Flow cytometry analysis and fluorescence microscope images of late G1 synchronized cells (upper panel) and asynchronous cells (lower panel) for *replicas* 1 (left) and 2 (right) are shown. B. MNase digestion profiles of replica 1 and replica 2 are displayed. The left panels show the size distribution of digested DNA molecules as measured by Bioanalyzer and the right panels show the agarose gel analysis of digestion products.

Replica 1 and 2 were both sequenced as paired-end (2x) reads, where both ends of DNA fragments were sequenced, assuming a minimal sequencing bias. The obtained sequencing reads (mean fold-coverage of 78x for individual experiments) were processed using the nucleR package (Materials&Methods,2.5.5)(Flores & Orozco 2011). To avoid gene average hindering effect in map comparisons, nucleosome profiles around TSSs were classified based on the nucleR score positioning of -1 nucleosome [(fuzzy (F), well-positioned (W) or missing (M)] and +1 nucleosome (F or W), and the width of NFR, defined as open (typically around 130 bp wide) or closed (around 30 bp wide) according to previously reported bimodal distributions (Zaugg & Luscombe 2011). For instance, a promoter with fuzzy -1 and +1 nucleosomes and an open NFR is called FoF (**Figure 32**). We were able to classify around 90% of yeast gene promoters into nucleosome architectures for both replicas. The remaining 10% could not be classified due to either low

coverage, undefined +1 or overlapping nucleosomes and were discarded in subsequent analyzes.

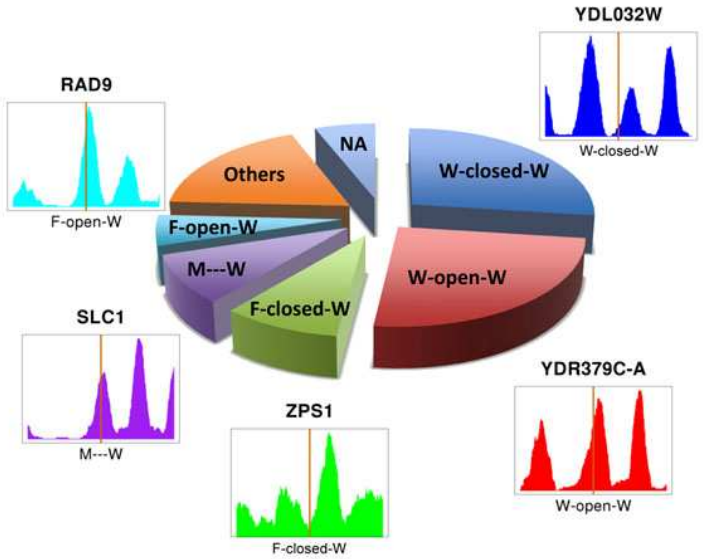


Figure 32: Gene clustering according to nucleosomal architecture at transcription start sites. Pie-chart shows the gene distribution for the most populated classes in the sample Replica 2 (2x). For every class, an example of the nucleosome coverage around the TSS of a representative gene is illustrated (window -300:300 from the TSS, marked in red)

Eventually, to minimize the noise arising from experimental procedures, we compared the nucleosome maps of these two synchronized replicas generated by paired-end sequencing. As shown in **Figure 33**, nucleosome patterns are quite similar in both replicas, dominated by WoW, WcW and a myriad of families characterized by a +1 W and -1 F/M. However, when individual genes were analyzed, clear differences arise between replicas (**Table 5**).

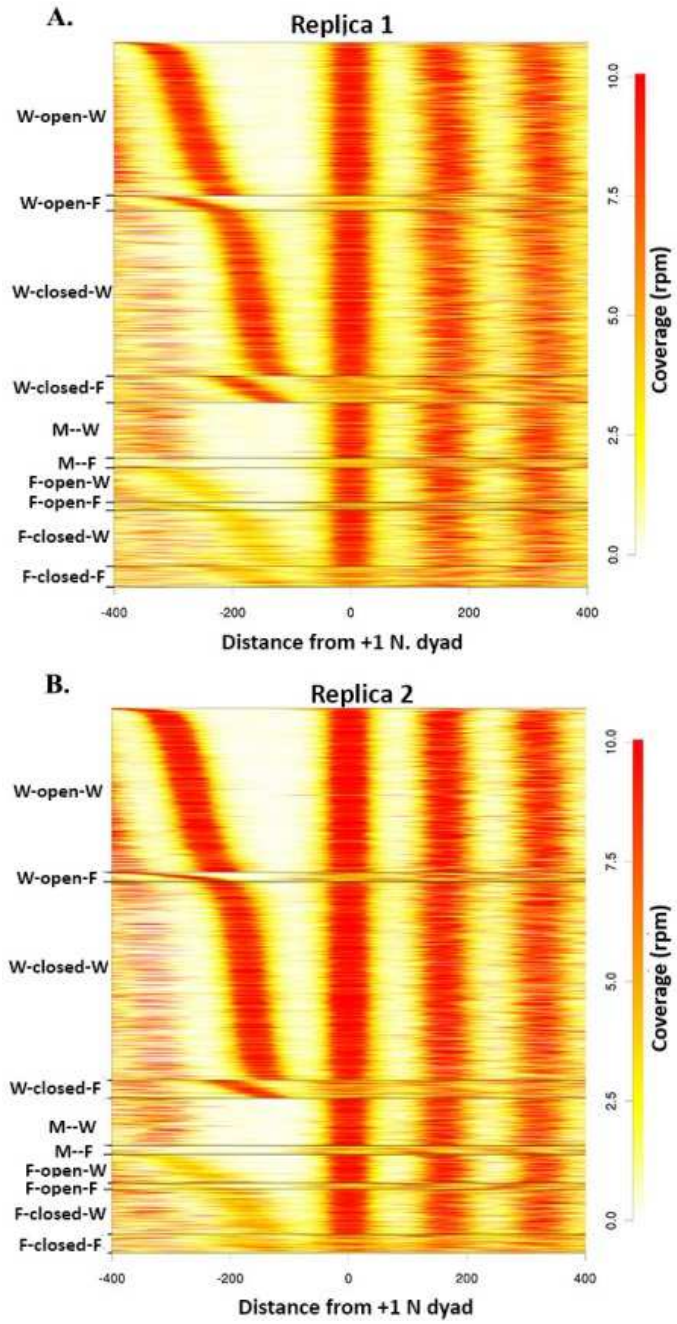


Figure 33: Nucleosome coverage and gene clustering. Heat maps show nucleosome occupancy around TSS in replica 1 (A) and replica 2 (B). Genes are clustered based on their nucleosome profile and their coverage is plotted taking +1 nucleosome dyad as ´0´. Colors represent the level of coverage (white: low, red: high).

Even though 90% of genes show similar coverage profiles, only around 60-67% of them maintain their -1/+1 nucleosomes and NFR classifications. The majority of the changes are subtle, usually only affecting one nucleosome position (e.g. W→F or F→M). Interestingly, when we focus on the variations, we observed that only 3% of genes show dramatic changes in NFR, in contrast to 6-7% that show different -1 and +1 nucleosome localizations (**Table 5**), indicating that NFR is more conserved and less prone to variations

	Same classification (%)				Variable classification (%)			
vs.	**Coverage**	**-1 Nuc.**	**+1 Nuc**	**NFR**	**Coverage**	**-1 Nuc.**	**+1 Nuc**	**NFR**
R1 R2	89.36	60.79	67.68	63.66	1.39	6.74	7.02	3.45
R1 As	76.55	61.73	71.1	66.98	3.41	12.82	12.24	4.81
R2 As	86.39	57.89	65.65	62.12	1.79	9.21	9.13	3.17
R1 Ov	83.92	59.26	71.85	66.53	1.36	12.55	12.81	4.81
R2 Ov	87.78	53.97	66.39	62.09	1.58	10.16	10.01	3.24
R1 Un	62.42	38.81	51.2	43.59	6.05	17.62	16.29	10.53
R2 Un	74.28	35.51	47.85	41.35	3.3	17.69	15.65	8.76
Ov Un	94.77	46.4	55.84	49.	0.28	10.77	10.92	8.17
As Ov	95.95	65.47	75.21	71.37	0.21	6.19	7.23	3.87
As Un	88.02	42.98	54.021	45.8	1.18	14.67	13.18	9.26

Table 5: Different pair-wise metrics of nucleosome similarity/dissimilarity. Coverage: A gene is considered as same (variable) if Pearson's correlation between two samples in -300:TSS:300 is greater than 0.7 (smaller than 0.5). +1/-1 Nucleosome: We consider a nucleosome in the same classification based on nucleR's classification variable if the absolute difference in nucleR's score is bigger than 0.25 points. NFR: we consider a gene stable if the classification of the NFR is the same; if the change in distance between -1/+1 nucleosomes is more than 100bp, it is a significant change. Percentages are relative to the total number of genes in the SacCer3 genome. Key: 2x – Paired End, R1 – Replica 1, R2 – Replica2, As – Asynchronous, Ov – Overdigested, Un – Underdigested.

than flanking nucleosome positions. These findings show that inter-replica variations are not negligible and point out that nucleosome positioning is intrinsically plastic and dynamic. In accordance with this observation, elastic energy models propose that a 10bp sliding of a nucleosome would face a maximum energy barrier (i.e. the difference between the best and worst wrapping configurations) of around 13 kcal/mol, which can be easily overcome, even in the absence of external effectors. On the other hand, at a larger scale, barriers of ~47 kcal/mol (**Figure 34A**) might contribute to nucleosome phasing. Local small energy fluctuations demonstrate that

nucleosomes tend to slide and change positions constantly, however at a global scale, where the energy barrier is much higher, their positioning is better phased. Therefore, well-positioned nucleosomes might not actually be tightly positioned in the absence of a high energy barrier such as nucleosome depletion signals, but largely fluctuate giving a general fuzziness signal. In order to explore whether fuzzy and well-positioned nucleosomes require distinct energy levels to be wrapped, we computed the deformation energy required to wrap each nucleosome (**Figure 34B**). As illustrated in **Figure 34B**, the deformation energies are indeed quite similar, around 200 kcal/mol for a 145 bp double-stranded DNA, revealing that fuzzy and well-positioned nucleosome formations are of equal difficulty. Fuzziness is then likely to be the default positioning state for nucleosomes in a random DNA fiber in the absence of additional factors, such as NFRs or protein effectors.

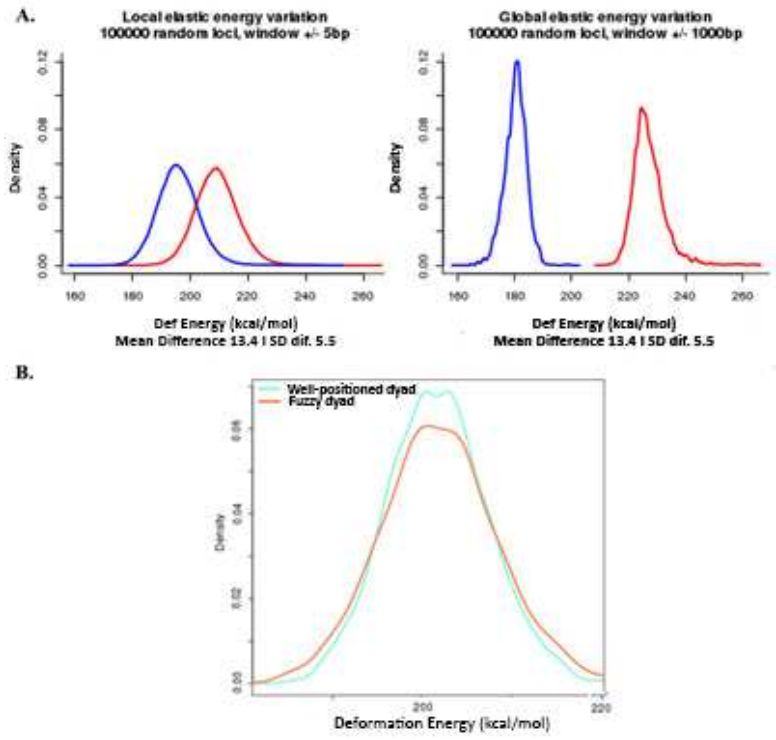


Figure 34: Energy barriers in nucleosome formation and deformation energy for nucleosome phasing. A. Local and global energy barriers are shown with minimum (blue) and maximum (red) values in a window of +/- 5bp (left) and +/- 1000bp right of 100000 random loci. B. Deformation energy around +/-5bp around the peak summit has been calculated for annotated -1/+1 nucleosomes. Mean value of every 10 possible combinations was used to account for local periodicity.

3.2 The Effect of Cell Diversity

In order to determine the variability derived from cell heterogeneity, we included an asynchronous yeast culture in our analysis, labeled as *asynchronous* sample (**Figure 31A**) and compared its nucleosome profile with synchronized *replicas* 1 and 2. Even though the most and the least predominant classes are the same, their distributions vary and *asynchronous* sample shows clear differences when compared to the synchronized *replicas* 1 and 2, as shown in **Figure 33**, **Figure 35,** Figure 36 and **Table 5**. The number of genes having WoW and WcW nucleosome classes decrease, while -1 F/M or +1 F nucleosome positions are more prevalent in asynchronous maps (proportion test p-value < $2.2 \cdot 10^{-16}$). The general increase in fuzzy nucleosomes suggests that cell cycle dependent chromatin reorganization around TSSs may be reflected as diffuse nucleosome signals in MNase-Seq experiments derived from asynchronous samples.

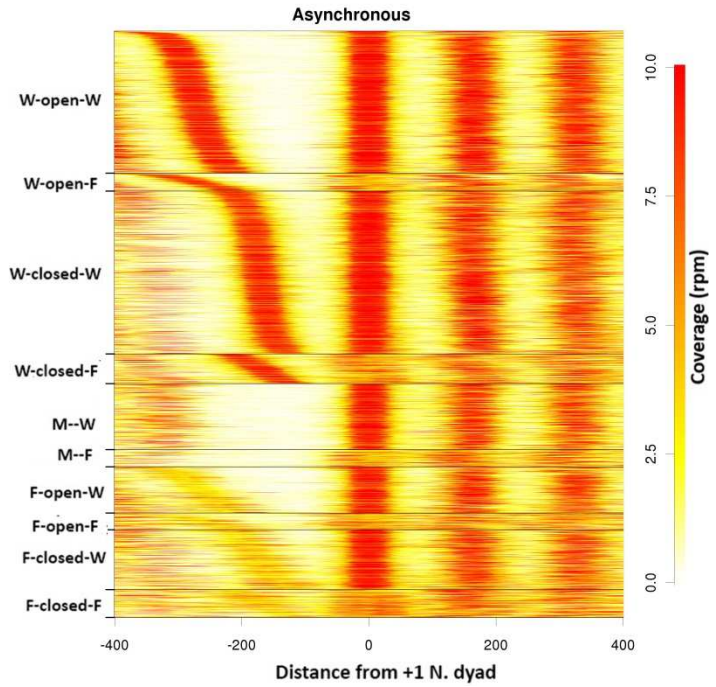


Figure 35: Nucleosome coverage and gene clustering. Heat map shows nucleosome occupancy around TSS in asynchronous sample. Genes are clustered based on their nucleosome profile and their coverage is plotted taking +1 nucleosome dyad as ´0´. Colors represent the level of coverage (white: low, red: high).

Moreover, analysis of individual genes showed a clearer impact of the variability caused by cell heterogeneity. The differences in nucleosome coverage profiles and architectures are much higher when the *asynchronous* sample is compared against biological replicas than between replicas (**Table 5, Figure 36**). The major changes seem to be in the phasing of -1 and +1 nucleosomes, which are fuzzier in *asynchronous* sample, as mentioned previously. In average, 475 genes with WoW or WcW classes in rep1/2 change their nucleosome configuration to more fuzzy structures in the asynchronous sample. As a result, the ratio of well-/non-well-positioned -1 nucleosomes decreases from 1.8-2.3 (*replicas* 1 and 2) to 1.4 (in *asynchronous*) (proportion test for *replica 1* p-value = $1.07 \cdot 10^{-10}$, for *replica 2* p-value < $2.2 \cdot 10^{-16}$), and similarly in case of +1 nucleosomes, dropping from 5.0-7.0 (*replicas* 1 and 2) to 3.7 (proportion test for *replica 1* p-value = $5 \cdot 10^{-11}$, for *replica 2* p-value < $2.2 \cdot 10^{-16}$). Conversely, NFR width between synchronized and asynchronous samples remains quite stable and shows similar changes as between biological replicas, suggesting that cell cycle-dependent chromatin rearrangements do not lead to massive nucleosome eviction around TSSs, which would dramatically alter NFR dimensions (proportion test for *replica 1* p-value = 0.22, for *replica 2* p-value = 0.09).

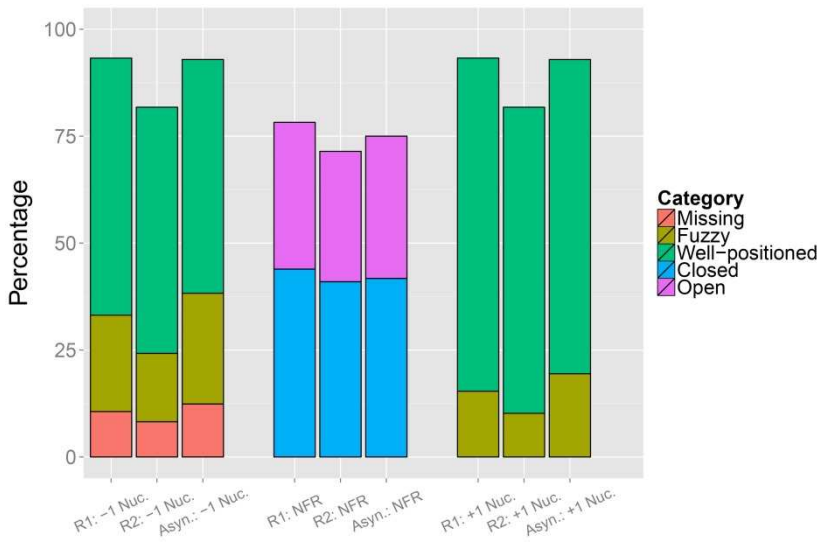


Figure 36: Distribution of -1/+1 nucleosomes and NFRs classification. -1 nucleosomes are classified as missing (M), fuzzy (F) or well-positioned (W) and +1 nucleosomes are either F or W. NFRs can have open or closed configuration depending on the NFR width. Key: R1 **– Replica 1,** R2 **– Replica2,** As **– Asynchronous.**

Overall, the comparison of cell cycle synchronized and *asynchronous* samples reveals that asynchronous experiments (Yuan et al. 2005; Lee et al. 2007; Mavrich et al. 2008; Kaplan et al. 2009) contain an additional source of noise due to the cell cycle-dependent nucleosome dynamics. Therefore, caution is necessary with maps derived from *asynchronous* samples (those typically available in the literature), since average maps can hinder diverse populations with completely distinct nucleosome architectures.

In order to test that, we have selected 211 genes, which show cell cycle periodicity and are highly active at G1 and 365 genes with transcription periodicity but highly active in the other stages (**Figure 37**). The analysis of their nucleosome profile around TSSs reveals that while synchronized samples display high +1 and -1 phasing in G1 active genes, the asynchronous sample shows a large perturbation with less defined peaks and linker regions. The same was also observed in the other cell cycle periodic genes. The differences in the asynchronous sample phasing confirm that different nucleosome architectures at each cell cycle stage introduce noise into average nucleosome profiles.

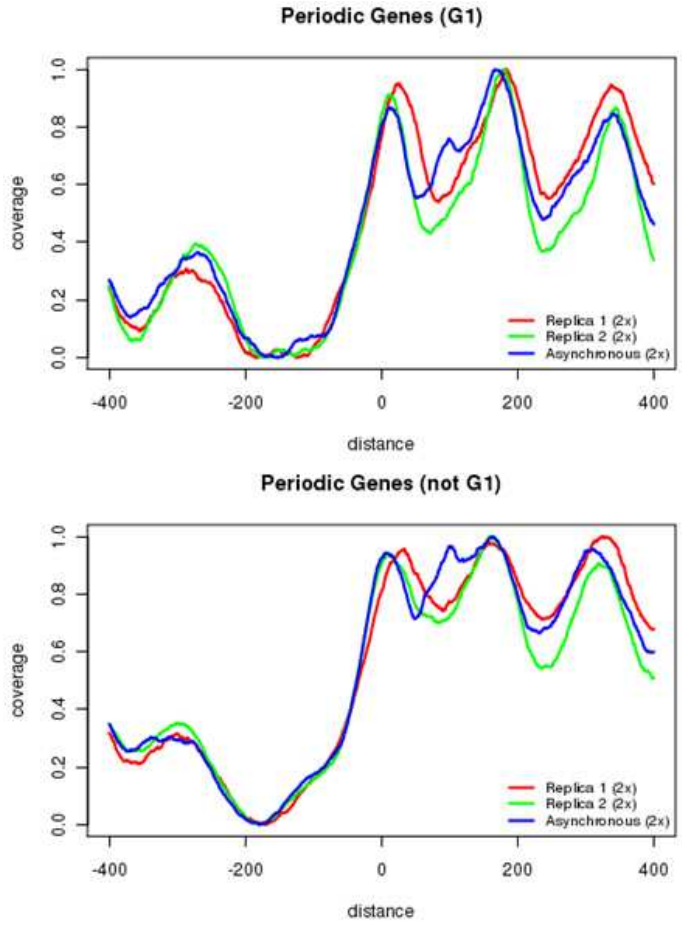


Figure 37: Effect of cell-cycle periodic genes in nucleosome map. Coverage of cell-cycle periodic genes is shown for G1 related genes (top, 211 genes) and in other stages (bottom, 365 genes).

In accordance, we observed that 105 genes displayed very similar coverage profiles around TSSs between biological replicas (Pearson's correlation > 0.7), but clearly differed with the *asynchronous* sample (Pearson's correlation < 0.5). Of note, although gene ontology (GO) and pathway annotation analyzes were not able to find any particular enrichment in this set of genes, 15 of them are annotated as cell cycle periodic genes in Cyclebase (Gauthier et al. 2010), which represents a small enrichment in cell-cycle related functions (from 9.8% in genomic mean to 14%; proportion test p-value=0.087).

3.3 The Effect of MNase Digestion

To investigate the bias of MNase digestion on the generation of nucleosome maps, we have used two additional MNase-Seq experiments derived from a G1-synchronized culture but treated under either more aggressive (*over-digested* sample*)* or milder (*under-digested* sample) MNase digestion conditions. As shown in the Bioanalyzer histograms (**Figure 38**), over-digestion of chromatin leads to the disappearance of the dinucleosome signal and to a broader mononucleosome peak that is shifted towards shorter fragments, probably caused by certain intra-nucleosomal cleavage. On the other hand, the *under-digested* sample exhibits well-defined mono-, di-, tri- and even tetra-nucleosomal peaks. The MNase-digested samples were then sequenced using paired-end technology and a similar data processing for a direct comparison with other replicas.

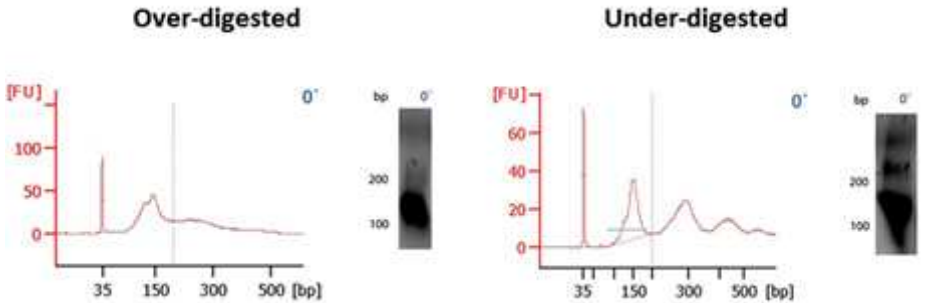


Figure 38: MNase digestion profiles of over- and under-digested samples. MNase digestion profiles of replica 1 and replica 2 are displayed. The left panels show the size distribution of digested DNA molecules as measured by Bioanalyzer and the right panels show the agarose gel analysis of digestion products.

Nucleosome architectures of the *over-digested* sample are well-defined with unambiguously assigned nucleosome families, similar to *replicas* 1 and 2. The analysis of nucleosome pattern distributions reveals clear differences between the *over-digested* sample and *replicas* 1 and 2 (**Figure 33** & **Figure 39**). In over-digested chromatin, the prevalence of canonical nucleosome classes (i.e. WoW and WcW) decreases, while fuzzy -1 nucleosomes are enriched. Explicitly, around 800 genes change their nucleosome patterns from WoW/WcW to a fuzzier configuration in the *over-digested* sample (proportion test p-value < $2.2 \cdot 10^{-16}$). The number of missing -1 nucleosomes increases from 711-510 (*replica* 1 and 2) to 1154 (proportion test p-value < $2.2 \cdot 10^{-16}$). Overall, these findings suggest that excessive MNase digestion can

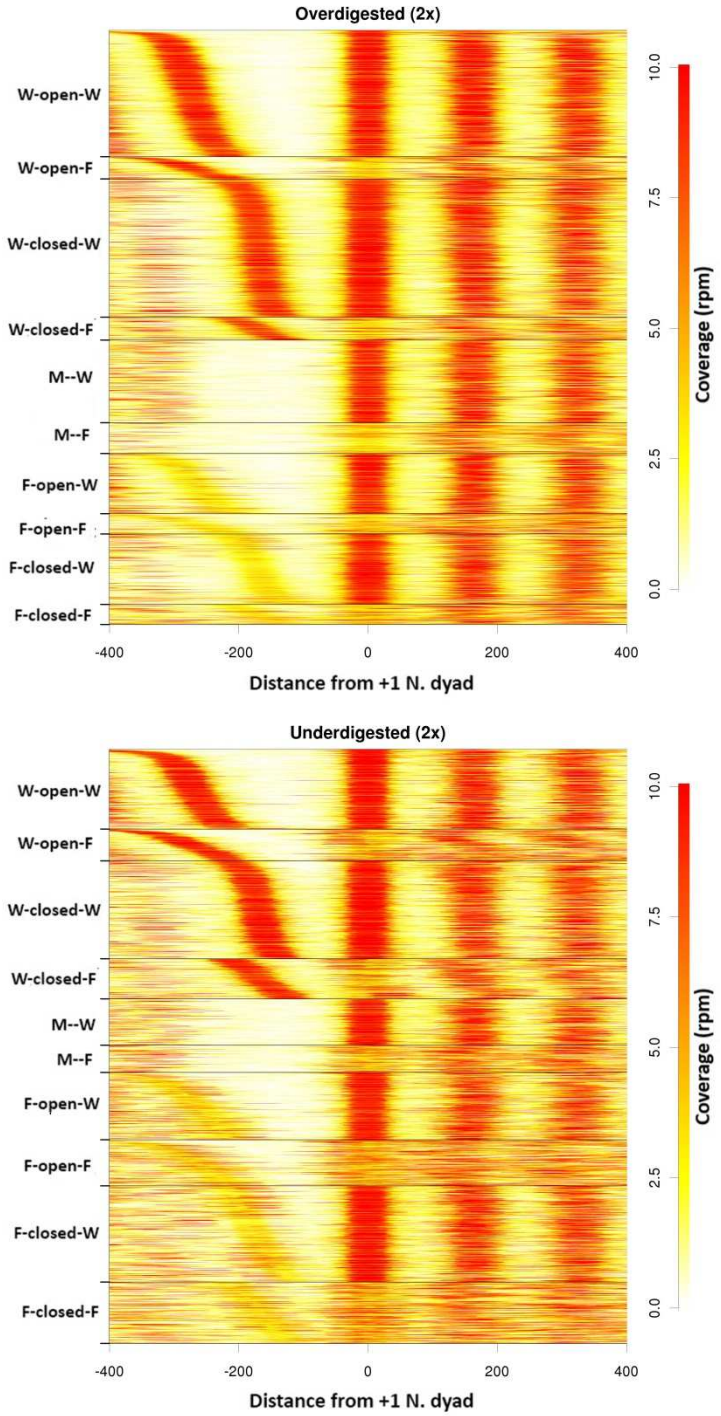


Figure 39: Over-digested & under-digested chromatin nucleosome coverage. Heat map shows nucleosome occupancy around TSS in *over-digested* and *under-digested samples.*

lead to partial degradation of some well-positioned nucleosomes, resulting in fuzzier nucleosome peaks or even to the complete disassociation of unstable nucleosomes, leading to loss of certain nucleosome signals. The effect of excessive digestion in nucleosome phasing is further confirmed with a higher mean deviation of the nucleosome dyad position in the *over-digested* sample (Kolmogorov-Smirnov test p-value = $2 \cdot 10^{-6}$), indicating that excessive MNase digestion leads to random intra-nucleosomal cleavage that is also reflected as fuzzy nucleosome signals (**Figure 40**).

We have further explored the impact of over-digestion on nucleosomal architectures by the analysis of individual genes (**Table 5**). While only 6-7% of the genes show clearly different -1 and +1 nucleosome annotations between replicas, up to 10-13% show variability with respect to *over-digested* sample. In contrast, NFR width seems to be very resistant to digestion conditions, pointing out that a more aggressive digestion mostly results in partial degradation of nucleosomes but rarely in their complete eviction around TSSs (**Table 5**).

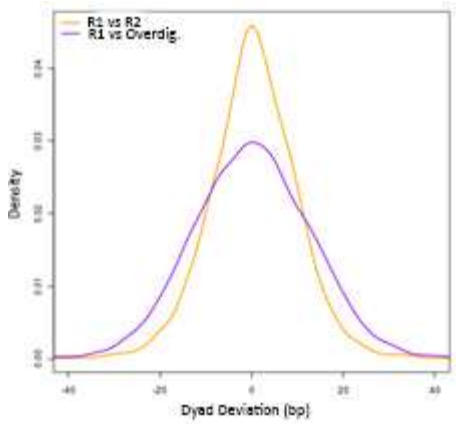


Figure 40: Comparison of dyad deviation in different MNase digestion conditions. Dyad distances of annotated -1/+1 nucleosomes (coverage peak summits) have been calculated between biological replicates and over-digested sample. Absolute mean deviation between Rep1 and Rep2 is 14.25 bp,, while it is 18.75bp (+4.5bp) between Rep1 and over-digested sample.

On the other hand, the analysis of *under-digested* sample revealed that only around 69% of TSSs could be classified into nucleosome families, whereas it was possible to classify 86.5 ± 3.7% of TSSs in previous experiments. The enrichment of depleted areas might account for the under-representation of longer fragments in the sequencing reactions, since it is well established that

deep sequencing favors the amplification of short over long fragments (Dabney & Meyer 2012). Interestingly, the uncovered regions specific to the under-digested sample are distributed over the entire genome without any significant enrichment according to GO analysis. Yet, they show a clear preference for AT-rich segments (3.38 higher fold) and intergenic regions (15% enrichment over background, simulated p-value < 10^{-5}). Despite the poor ability of sequencing procedures for longer fragments, we were able to recover several long reads analogous to dinucleosomal signals, among which 3% are longer than 300 bp in the *under-digested* sample. In contrast, we only rescued 0.4% of these long reads in the *over-digested* sample. For that reason, we studied whether the dinucleosome signals might introduce another source of noise in the nucleosome maps, since nucleosome calling algorithms align the fragments based on their middle position assuming that will correspond to the nucleosome dyad. However, this is not the case for dinucleosome fragments, leading to misaligning of the dinucleosomes. Therefore, we compared coverage profiles of short-, mid- or long-sized fragments in *over-* and *under-digested* sample. The selections of read lengths are shown in **Figure 41** on the right plots. We have observed that while short- and mid-sized fragments yield to mononucleosome signals, long-sized fragments lead to a counter-phase location with respect to mononucleosome signals. The counter-phasing is more explicit in under-digestion, which contains more dinucleosome derived signals, as noted previously. Therefore, longer fragments lead to a higher noise in terms of linker length and nucleosome phasing, which is not typically considered in nucleosome maps (**Figure 41**).

Overall, our observations show that MNase digestion levels may strongly bias nucleosome maps by intra-nucleosomal cleavage or longer inter-nucleosomal fragments. Therefore, caution is necessary, when maps obtained under different digestion conditions are compared. This warning is especially important since MNase is an enzyme whose activity is not always easy to control.

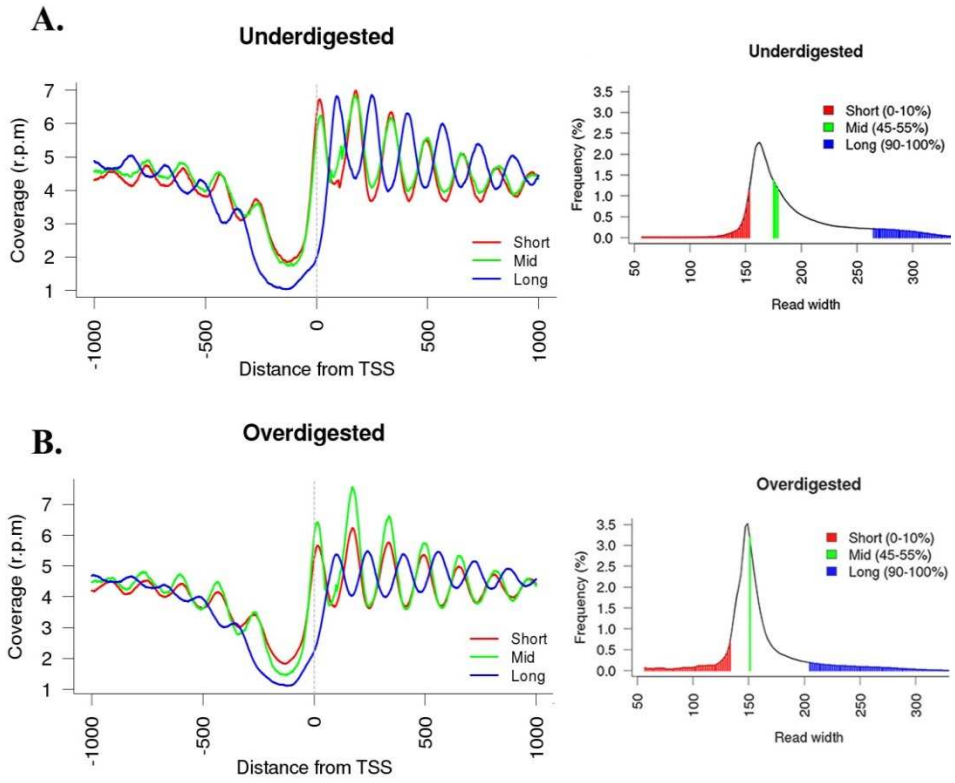


Figure 41: Effect of variable read length on map coverage. Normalized coverage profiles of trimmed reads around TSSs is shown on the left side and coverage distribution of short, mid and long reads is shown on the right side for A. *under-digested* and B. *over-digested samples*.

3.4 Underlying Factors in Nucleosome Positioning

For further analysis, in order to minimize the variability and noise in the nucleosome maps that we have detected, we have chosen a robust set of nucleosome profiles, which show a correlation greater than 0.7 and display the same nucleosome architecture in biological replicas. This set comprising 3096 genes represents the well-conserved nucleosome architectures in late G1 cell cycle phase and reveals that WcW (1306 genes) and WoW (1164 genes) are the most pervasive classes, followed by M-W (263 genes) and FcF (155 genes) classes. We used this set of profiles to evaluate different predictive models.

We compared *in vivo* nucleosome profiles of genes having WcW and WoW pattern with a simple statistical model that locates nucleosomes with

decreasing phasing every 161-165 bp (147 + 14-18 bp long linker; **Figure 42A**). Once NFR is defined, the model is able to predict the majority of nucleosome positioning around TSSs, supporting the barrier model as the major nucleosome positioning determinants and thus NFR as the barrier. To analyze what determines NFR position, we studied transcription factor binding sites (TFBSs) and compute the deformation energy required to wrap DNA around the histone core around TSSs (**Figure 42B**). NFRs at WoW classes display larger deformation energies as a consequence of the DNA properties at these regions, indicating that physical properties can define the boundaries of the NFRs in this family. However the lower values of deformation energy in the middle of the NFRs also signal a well-positioned nucleosome at that area, where predictive TFBS signal is very high. This observation shows that while the NFR centers intrinsically allow nucleosome formation, the competition with transcription factors avoids the potential binding of the nucleosome in this region. Clearly, TF binding is then crucial in determining the integrity of NFRs and hence the phasing of the nucleosome arrays in WoW architectures. This synergetic effect of physical properties and TFBS on nucleosome phasing is also clear in the WcW family, where the region around TSSs is marked by an unusual profile of physical properties and a distinct pattern of TFBSs. The strong +1 nucleosome signal fits perfectly in a region of low cost for wrapping DNA around a nucleosome and depleted in TFBS. Similarly, NFRs display higher deformation energy and TFBS density that prevent nucleosomes to occlude NFRs.

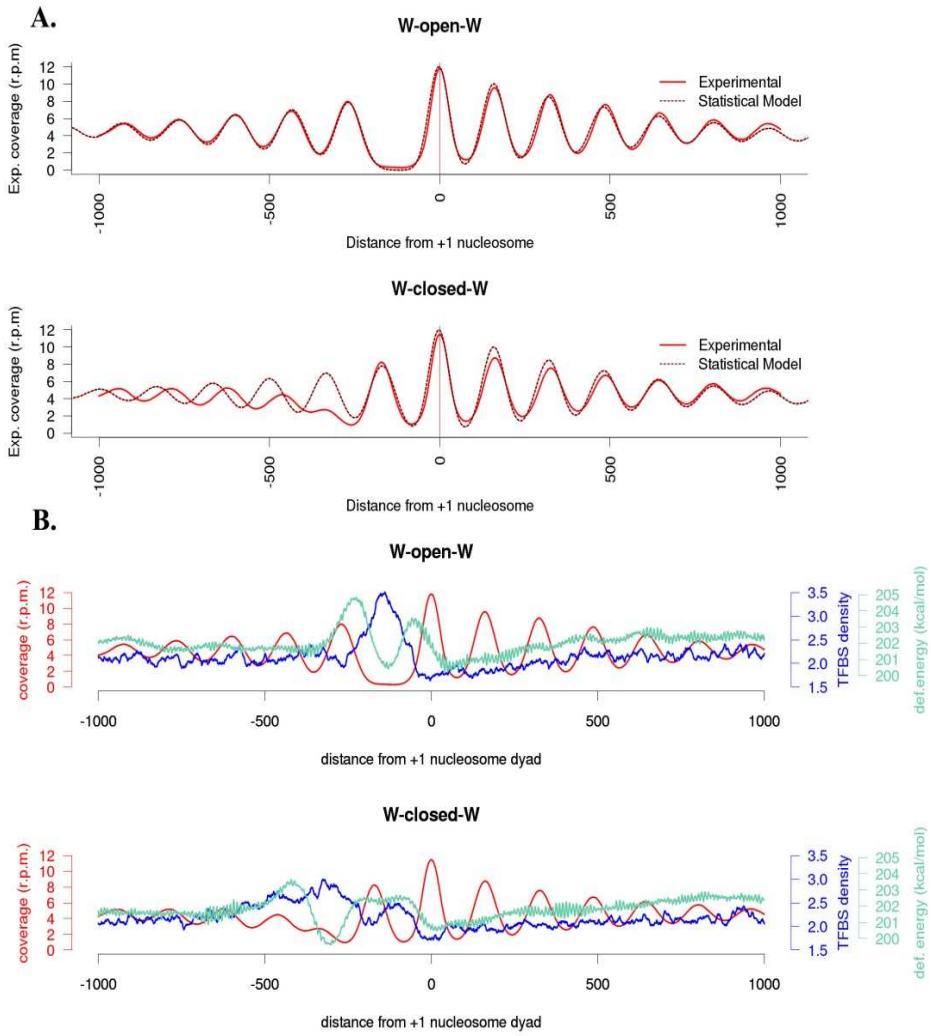


Figure 42: Statistical positioning and intrinsic DNA energetic barriers. A. Average experimental nucleosome coverage from WoW (top) and WcW (bottom) patterns in *replica 1* are compared against a nucleosome positioning statistical model. B. The experimental coverage of WoW (top) and WcW (bottom) classes (red) are overlapped with deformation energy (cyan) and predictive TFBS (blue) around TSSs.

This study has been recently accepted to NAR (NAR-00256-X-2014.R1)

4 Chromatin Dynamics throughout Cell Cycle

In eukaryotic cells, chromatin organization varies as cell cycle progresses, especially at S phase, when DNA replication takes place, and at M phase, when chromosomes condense. Additionally, the expression of approximately 800 cell cycle-regulated genes that varies along cell cycle should be partially modulated by chromatin structure at gene promoters. Therefore, in order to study the relationship between chromatin dynamics and gene expression and to study cell-cycle dependent variations in chromatin properties we have analyzed nucleosome positioning along cell cycle using synchronized populations of *S. cerevisiae.*

4.1 Cell Synchronization and Determination of Cell Cycle Duration

The cells were arrested at late G1 by the alpha-factor mating pheromone (Materials&Methods 2.2.1). Upon treatment, the cell morphology was inspected every 15' until all the cells were un-budded and had shmoo shape, typically after 2 h. The arrested cells were then released into fresh YPD medium, and synchronized cell populations were collected at 10-15' intervals. Cell cycle phase duration was determined based on the cell morphology by fluorescence microscopy**. Figure 43** shows fluorescence microscopy images of BY4741 cells, which were collected at indicated time points and stained with HOECHST. At time 0', the cells are in alpha factor-containing medium and they are all un-budded with shmoo shape due to the response to alpha factor. At 15', the buds start to emerge, indicating S phase entry. The smaller buds at 25' and 35' are indication of S phase, while larger budded cells at 45' are in G2 phase. At 60' the cells are already at G2/M transition, since the nuclear migration takes place. Completion of anaphase is seen at 70' with the presence of divided nuclei. Cytokinesis takes place at 85'. Overall, the estimated cell cycle duration of these cells is 90-100'.

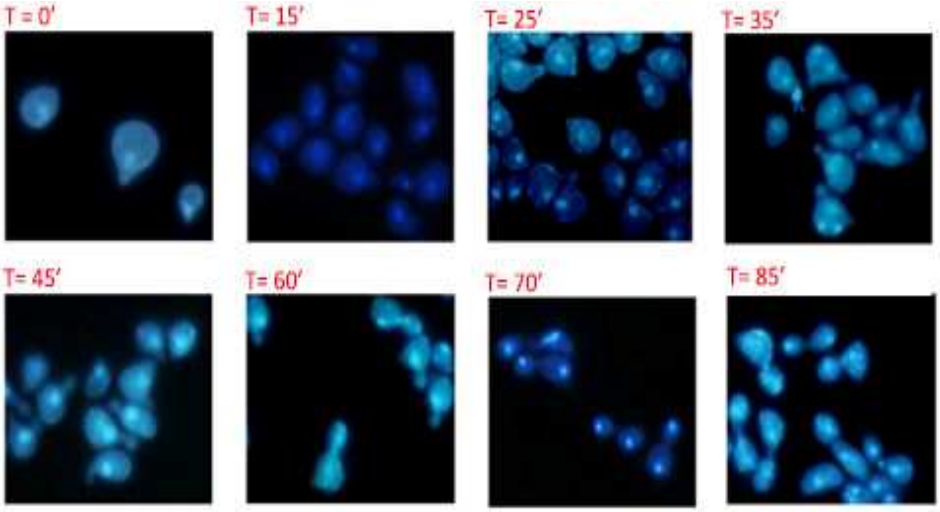


Figure 43: Fluorescence microscopy images of BY4741 cells. The images were captured at the indicated times (min). The total cell cycle was estimate to be around 90'.

4.2 Cell Cycle Synchrony Monitorization

Once every cell cycle phase duration was estimated, samples were collected at every stage. For each experiment, cell population synchrony was monitored and verified by three approaches: flow cytometry (FACS), fluorescence microscopy and budding index calculation, as described in Materials&Methods 2.2.2 and shown in **Figure 44**. FACS measures DNA content by propidium iodide (PI) staining, which binds to DNA in a stoichiometric way allowing quantification of DNA amount in every cell. FACS analysis shows that as the cells proceed to S phase (30'), DNA amount increases due to DNA replication and reaches its maximum at M phase (70'), before cytokinesis. Once the cytokinesis takes place, DNA amount per cell decreases again (90') and some cells already start the new cell cycle. Furthermore, the cycle progress is monitored based on the cell morphology by fluorescence microscopy, as explained in the previous section. Finally, budding index gives more information about bud emergence. As seen in **Figure 44C**, the buds already start emerging after 10' and at 30' half of the populations is budded, indicating S phase entry. It is also worth noting that the population starts losing the synchrony already after 70', as observed by FACS and microscopy. Even though the majority of the cells are at M phase, some of the cells proceed to cytokinesis and some of them still remain at G2

phase. Therefore, samples are not collected after this point, since the loss of synchrony might interfere with the interpretation of the results.

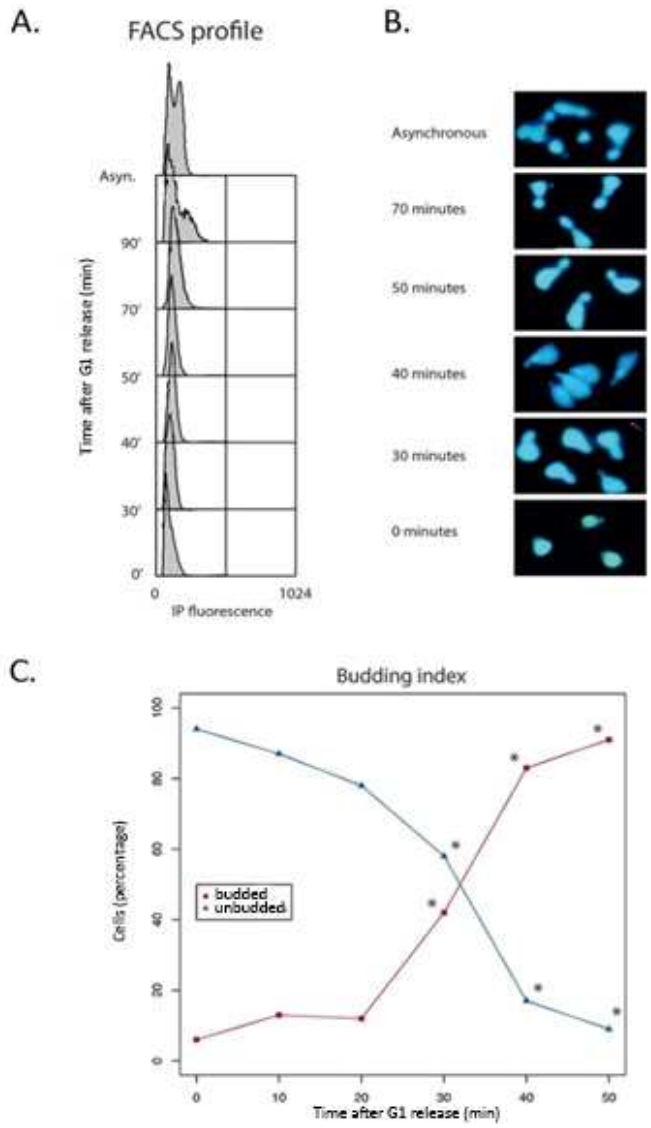


Figure 44: Monitoring cell cycle progress of BY4741 cells, by flow cytometry (A), fluorescence microscopy (B) and budding index (C).

4.3 Chromatin Dynamics along Cell Cycle

4.3.1 Global Nucleosome Dynamics

We studied chromatin dynamics along cell cycle by MNase digestion assays and MNase-Seq analysis.

To see the general effect of cell cycle in chromatin dynamics, I first performed MNase digestion assays by collecting 50 ml synchronized cultures at each cell cycle phase after alpha factor release. Once the cell cycle stages were verified by fluorescence microscopy (**Figure 45A**), cells were spheroplasted and the chromatin extracts were digested with 0.1 U of MNase. The deproteinized digested samples were run on 2% agarose gel to inspect the nucleosomal DNA patterns, as shown in **Figure 45B**.

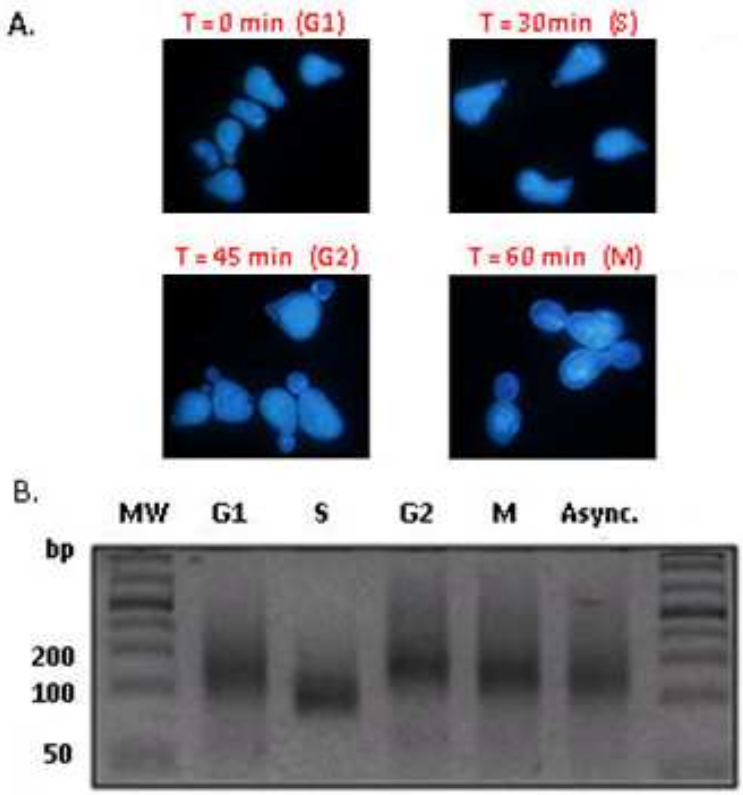


Figure 45: MNase digestion pattern along cell cycle. A. Yeast cells at G1, S, G2 and M phases are monitored using Hoechst stain by fluorescence microscopy. B. Samples collected along cell cycle and asynchronous sample are digested by 0.1 U MNase (25', 37°C) and fractionated on a 2% agarose gel.

Asynchronous sample, G1, G2 and M phase samples yield mostly a mononucleosome band of 150 bp and partially longer fragments up to 300-400 bp, corresponding to di- and tri-nucleosomes. However, the MNase digestion of S phase chromatin displays a distinct pattern, where fragments are more uniform and shorter, resulting in a single band around 100 bp, indicating that S phase chromatin has a distinct conformation and is more sensitive to MNase.

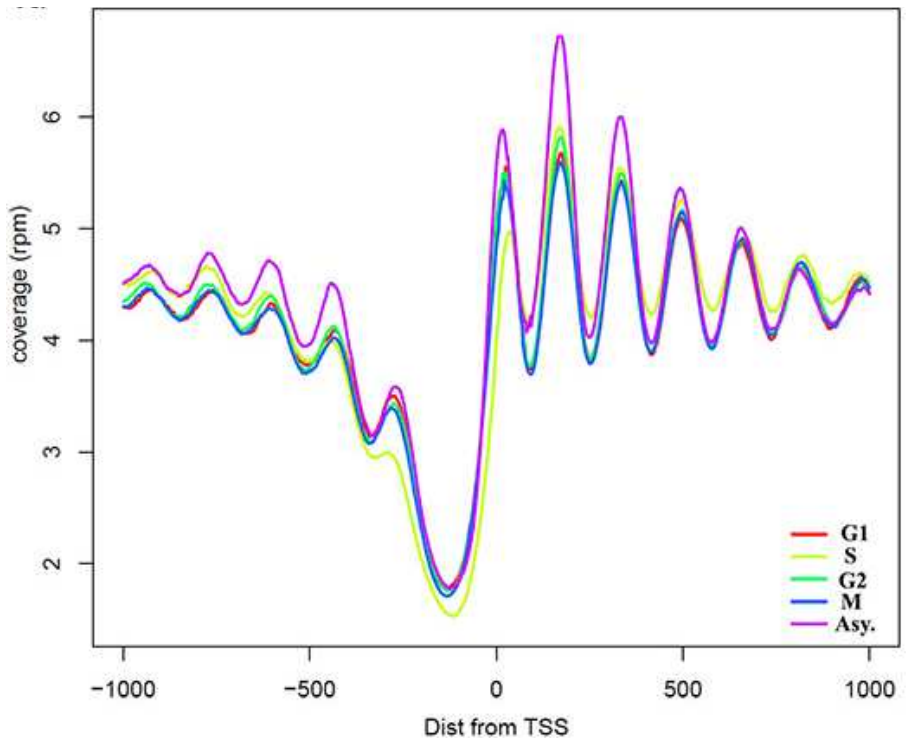


Figure 46: Comparison of nucleosome profiles along cell cycle. A. Coverage maps per base pair were calculated and averaged across all yeast genome around TSSs.

To analyze the chromatin dynamics along cell cycle at the mononucleosome level, the DNA samples were single-end sequenced on an Illumina/Solexa Genome Analyzer (GA) IIx with 38 bp read length. The obtained reads were aligned on the reference yeast genome (Saccer 3, 2011) with Bowtie software and processed with R/Bioconductor package (nucleR) to generate nucleosome maps. The general nucleosome pattern is mainly conserved and all cell cycle stages display a canonical nucleosome organization around TSSs, where NFRs are located upstream TSSs and flanked by phased nucleosomes (**Figure 46**). However, there are some notable differences in nucleosome occupancy and phasing in S phase chromatin. -1 and +1 nucleosomes have

lower occupancy and 5' NFRs are slightly deeper and wider. Moreover, S phase chromatin displays more fuzziness in -1 nucleosome and downstream nucleosomes, which are marked by less-well defined linker regions and nucleosome peaks.

In order to test whether S phase fuzziness is only specific to promoter regions, we have calculated the total number of well-positioned and fuzzy nucleosomes at each cell cycle stage (**Figure 47**). While 30-35% of annotated nucleosomes are fuzzy at G1, G2 and M phases, the fuzzy nucleosomes at S phase comprise almost 50% of the identified nucleosomes, indicating global nucleosome fuzziness at S phase.

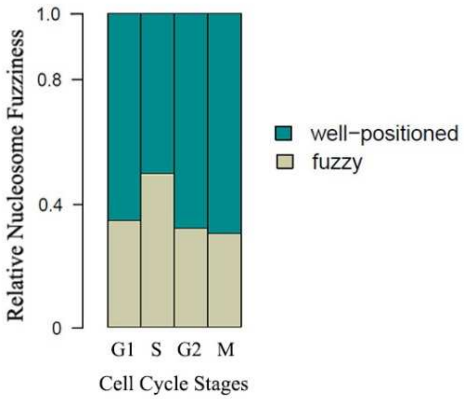


Figure 47: Relative nucleosome fuzziness along cell cycle. Total number of well-positioned and fuzzy nucleosomes was calculated along the genome and their relative values were compared at each cell cycle stages.

Even though single-end sequencing does not enable a more detailed analysis, our results show that S phase has a higher degree of chromatin disorganization, indicating that S phase may encompass a different nucleosome-based mechanism of regulation in a DNA replication dependent manner, in line with previous suggestions (Hogan et al. 2006).

4.3.2 Nucleosome Dynamics at Particular Cell Cycle Points

In an attempt to examine whether higher order chromatin structure displays increased sensitivity to MNase digestion throughout the entire S phase, two additional samples were collected at S phase with 3' intervals. Formaldehyde cross-linked chromatin was partially digested with increasing amount of MNase. The pattern of MNase digestion shows that samples at 0' and 60', corresponding to G1 and M phases, display an overall more intense

nucleosome ladder, while the samples at 27', 30', 33' and 45', corresponding to S and G2 phases, exhibit an increased sensitivity to MNase digestion (**Figure 48A**). To analyze MNase cleavage pattern in details, the ratio between mononucleosomes and di-, tri- or tetra-nucleosomes intensities were calculated using IMAGEJ software. Remarkably, the ratio of mono- to di-nucleosomes remains the same among the samples, whereas the ratio to tri- and tetra-nucleosomes exhibits high variations along cell cycle. 60' sample has the lowest mono- to tri- or tetra-nucleosome ratio, indicating lower MNase accessibility. 0' and 45' samples, G1 phase and G2/M transition, display a very similar MNase cleavage pattern with relatively higher sensitivity to MNase digestion. On the other hand, the samples collected throughout S phase, 27', 30', 33' samples, show an increased level of MNase digestion, confirming the previous results (**Figure 45**). Interestingly, the samples collected at S phase do not exhibit the same level of sensitivity to MNase. 30' sample particularly has the highest mono- to tri- and tetra-nucleosome ratio. These ratios decrease gradually before and after 30', demonstrating a step-wise chromatin compaction along cell cycle with peak sensitivity at one particular time point in S phase.

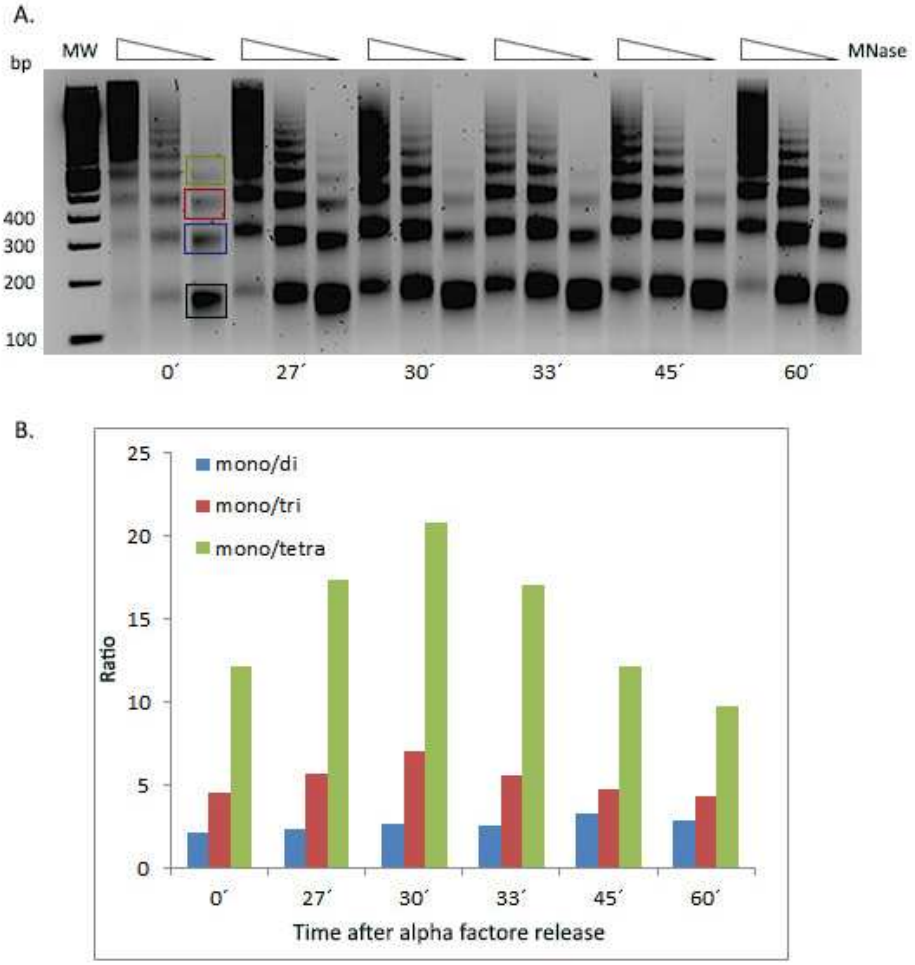


Figure 48: Chromatin sensitivity to MNase digestion along cell cycle. A. Chromatin isolated at indicated time points after alpha factor release was digested with 0.005, 0.01 and 0.025 U of MNase (25', 37°C) as shown by the triangles above the lanes. B. The bar plot shows the ratio between mononucleosomes and di-, tri- or tetra-nucleosomes, calculated using IMAGEJ, of cell cycle samples.

Since partial MNase digestion does not result in optimal fragments for sequencing, we performed similar experiment by collecting samples at 0', 60' and between 28' and 45' every 2-3' intervals. Nuclei were then digested with excessive MNase yielding mostly mononucleosomes (**Figure 49A**). Even though digested DNA samples yield mostly mononucleosome fragments, the digestion patterns show some variations along the cell cycle. Mono-, di- and even tri-nucleosome bands are visible in the asynchronous, 36' and 39' samples, which indicates less digested chromatin; while 0', 28', 33', 42' and 60' samples display very faint dinucleosome and strong mononucleosome

bands. On the other hand, 31’ and 45’ samples display the highest sensitivity to MNase digestion. 31’ mononucleosome band is barely visible and that of 45’ sample is noticeably weak. 31’ corresponds to mid-S phase in *S. cerevisiae*, as defined from fluorescence microscopy images. 45’ corresponds to G2/M transition. However, since daughter cells have a longer G1 phase, they enter to S phase relatively later than mother cells (Hartwell & Unger 1977). Hence, this 45’ time point might correspond to S phase of the daughter cells (Brewer et al. 1984). Interestingly, the particular high sensitivity to MNase is not conserved throughout entire S phase, but rather observed within a very short time period, similar to previous observations (**Figure 48**). Chromatin accessibility seems to quickly decrease after 2-3’, as the mononucleosome bands become more visible (**Figure 49A**). Similar chromatin behavior is also observed for the 45’ sample, whereas chromatin at 42’ is more resistant to MNase digestion.

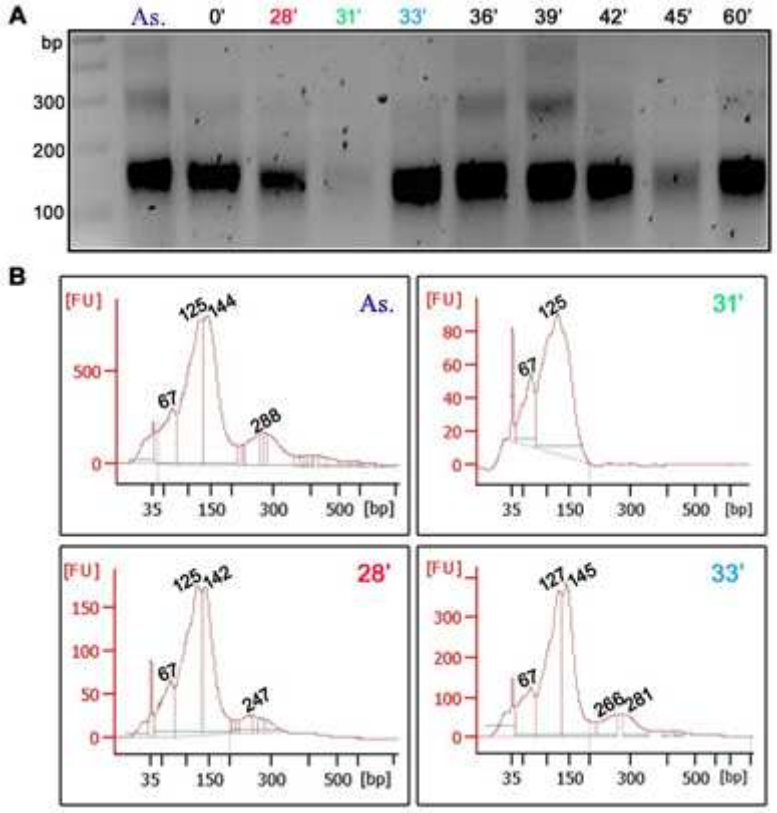


Figure 49: MNase cleavage assay along cell cycle at shorter intervals. A. Samples collected at indicated time points after alpha factor release and asynchronous sample are digested by 0.1 U MNase (25’, 37°C) and fractionated on 2% agarose gel. B. Bioanalyzer profiles of the samples show their MNase digestion pattern.

In order to further examine the differential fragmentation of chromatin at S phase, the samples were analyzed by microfluidics-based platform Bioanalyzer (**Figure 49B**). Asynchronous sample, samples at 28' and 33' yield a major peak around 145 bp, corresponding to mononucleosomes, and a second major peak around 125 bp, due to the digestion of nucleosome edges. These samples also display an additional peak around 295 bp, corresponding to dinucleosomes, and residual peaks at around 60 bp that might be assigned to either tetrasomes or other DNA-protein complexes. On the other hand, chromatin at 31' exhibits a broader mononucleosome peak that is shifted towards shorter fragments and the smaller peak around 67 bp is much higher compared to the other samples. Moreover, the dinucleosome signal is completely absent, as observed in **Figure 49A**. Interestingly, the high MNase sensibility can be recovered after 2-3'.

We repeated the same experiment three more times to ensure the reproducibility of our findings. Indeed, the highest sensitivity was observed around 30' in one sample and around 45-50' in another sample (Appendix 1), while the rest of the samples yielded less digested mono- and dinucleosome fragments. As observed in the three biological replicas, these chromatin changes are sudden and acute; thus it is very difficult to catch those points unless the chromatin accessibility is analyzed within short time intervals. This may probably explain why these findings were not observed in our previous cell cycle experiments, carried out at longer time intervals.

In order to more accurately study the global and local nucleosome dynamics along cell cycle, the samples at 0', 28', 31', 33', 36', 39', 45' and 60' and asynchronous sample were paired-end sequenced on an Illumina/Solexa Genome Analyzer (GA) IIx with 100 bp read length, and similarly processed as for single-end sequencing data. The average nucleosome profile around TSSs is quite similar among the samples. We could not observe any dramatic changes among the samples at 0', 28', 31', 33', 45' and 60', even though the bioanalyzer profiles show higher amount of smaller fragments for 31' and 45' samples (**Figure 49B**). However, chromatin at time points 0, 36 and 39' shows higher occupancy and phasing at -1 and +1 nucleosomes (Appendix 2).

Of note, since the read length is now 100bp, those fragments that are smaller than 100 bp will not be sequenced, which then impedes the detailed study of the smaller fragments at 31' and 45' For that reason, we have focused our analysis on 4 samples: 0', 33', 39' and 60', representing G1, S, G2

and M phases. We have verified the phase status based on microscopy and gene expression data. Interestingly, lower nucleosome occupancies are notable at S and M phases, especially at -2, -1 and +1 nucleosomes, and 5' NFRs are slightly longer and wider at S and M phases (**Figure 50**). Moreover, the linker regions are better defined and deeper at G1 and G2 phases, indicating higher nucleosome phasing compared to S and M phases.

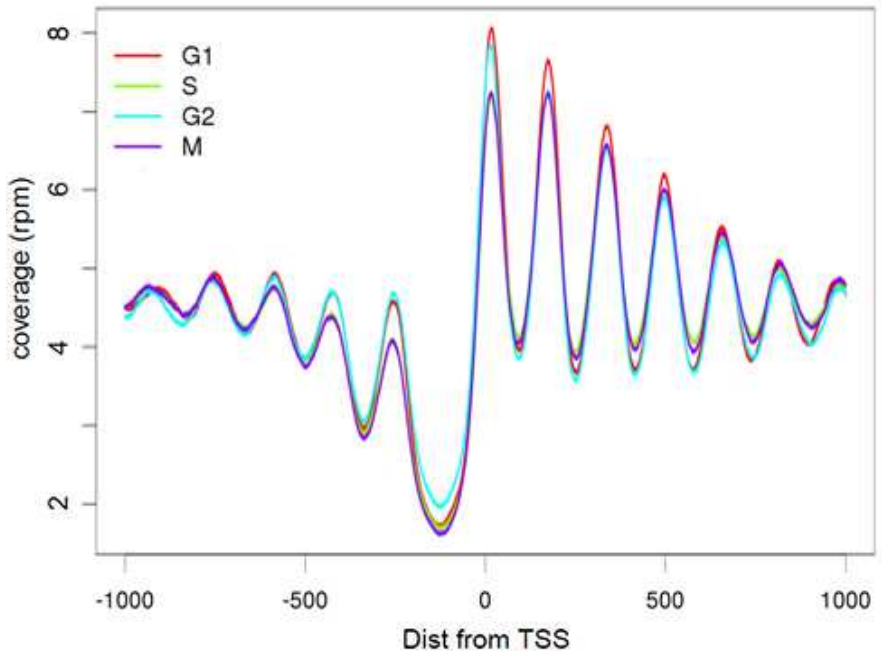
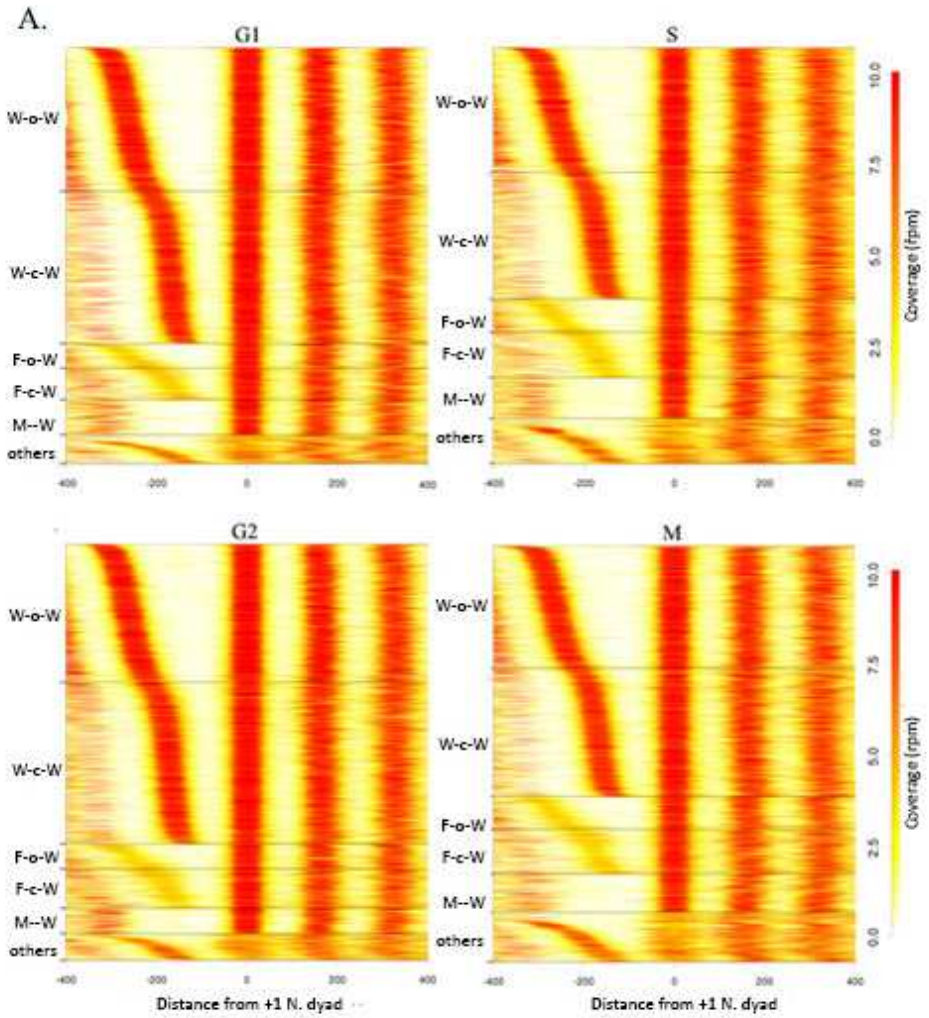


Figure 50: Comparison of nucleosome profiles along cell cycle. Coverage maps per base pair were calculated and averaged across all yeast genome around TSSs at G1 (red), S (green), G2 (blue) and M (purple) phases.

To avoid gene average hindering effect, nucleosome profiles around TSSs have been classified based on the nucleR score positioning of -1 nucleosome [(fuzzy (F), well-positioned (W) or missing (M)] and +1 nucleosome (F or W)], and the width of NFR, defined as open (typically around 130 bp wide) or closed (around 30 bp wide) (Materials&Methods,2.5.6). We have generated heat maps of nucleosome profiles around TSSs using 3279 genes that are assigned to one of the classes and with available gene expression data (**Figure 51A**). In each cell cycle stage, nucleosome profiles are dominated by WoW, WcW, whereas nucleosome classes with +1 W and -1 F/M are minor. However, when the distribution of classes are compared, we observe that at S and M phases there is a decrease in the number of the genes in WoW and

WcW classes, while the number of genes in classes containing fuzzy or missing -1 nucleosomes are higher.

A.



B.

	-1 Nucleosome			NFR		+1 Nucleosome	
Stage	F-*	W-*	M-*	*-open-*	*-closed-*	*-F	*-W
G1	16%	74%	10%	48%	42%	7%	93%
S	23%	66%	12%	47%	42%	11%	89%
G2	18%	75%	7%	52%	41%	6%	94%
M	22%	66%	11%	48%	41%	12%	88%

Figure 51: Comparison of nucleosome profile around TSSs along cell cycle. A. Heat maps show nucleosome occupancy around TSS in G1, S, G2 and M phases. Genes are clustered based on their nucleosome profile and their coverage is plotted taking +1 nucleosome dyad as ´0´. Colors represent the level of coverage (white: low, red: high). B. Distribution of classifications for -1 nucleosome (left), NFR (center) and +1 nucleosome

As a next step, we separately compared the distribution of fuzzy and well-positioned -1/+1 nucleosomes and NFR widths along cell cycle. **Figure 51B** demonstrates that at G1 and G2 phases, 74-75% of the genes have well-positioned -1 nucleosomes and 93-94% display well-positioning at +1 nucleosomes. On the other hand, at S and M phases the number of genes having well-positioned nucleosomes dropped by 10% for -1 and 5% for +1 nucleosomes. Interestingly, the genes having open or closed NFR status do not seem to change along cell cycle, showing that the alterations in nucleosome phasing do not alter the distance between them.

We have further compared the global fuzziness at each cell cycle phase. As seen in **Figure 52A**., S and M phase chromatin does contain more fuzzy nucleosomes along the genome, while chromatin at G1 phase has the highest number of well-positioned nucleosomes, followed by G2 phase chromatin. In accordance with the well-positioned/ fuzzy nucleosome distributions, the correlation coefficient is high between G1 and G2 (0.89); while it is slightly lower between G1 and S or M phases (0.87 or 0.84). Notably, these pattern differences are not related to a significant change in overall gene expression based on microarray hybridization experiments, since it does not vary along cell cycle (please see section 4.4 for gene expression data) (Appendix 3).

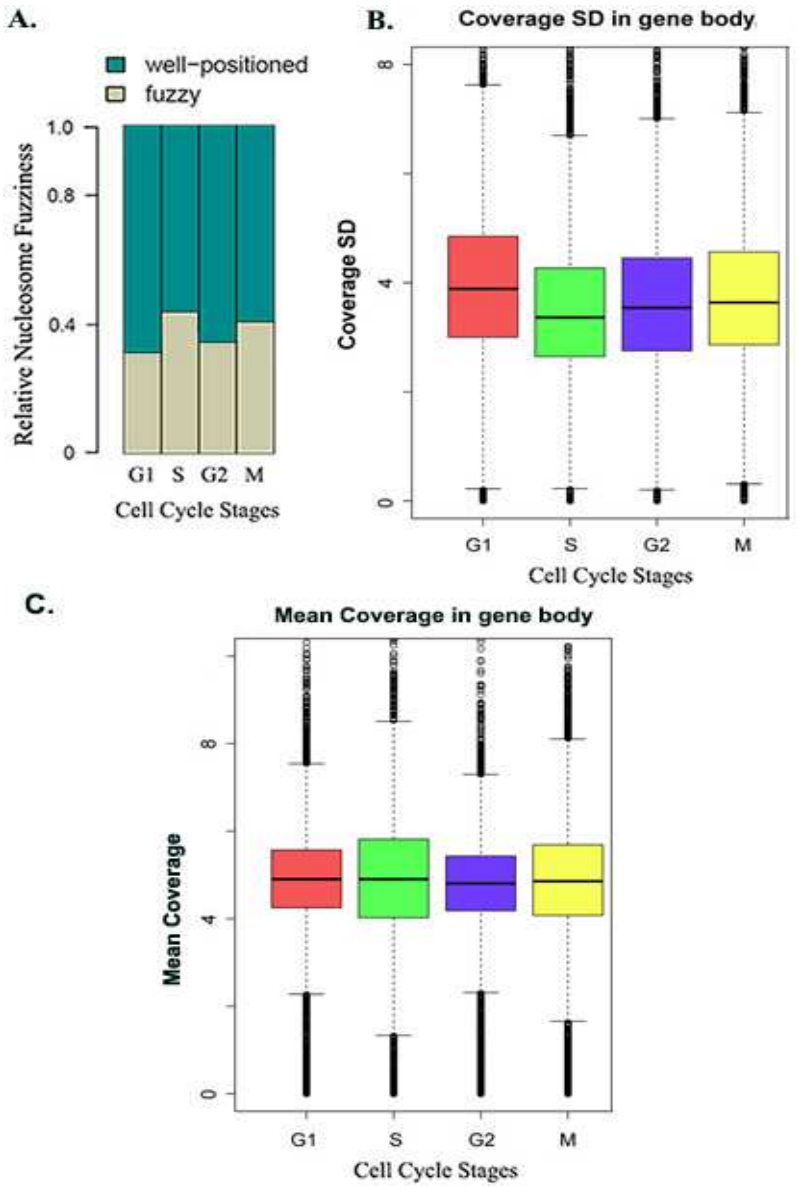


Figure 52: Global nucleosome phasing along cell cycle. A. Total number of well-positioned and fuzzy nucleosomes were calculated along the genome and their relative values were compared at each cell cycle stages. B. The boxplot illustrates standard deviation (SD) of the normalized coverage (reads per million) that is calculated at each cell cycle and compared at the gene bodies. C. Boxplot shows the mean of the normalized coverage (in reads per million) at each stage within the gene bodies.

We have analyzed the standard deviation (SD) of the normalized nucleosome coverage along gene body at each cell cycle stage as a measure of nucleosome occupancy variation (**Figure 52B**). A high standard deviation

implies sharper peaks, while low values mean that nucleosome depleted or enriched regions are not well-defined. Accordingly, coverage peaks at G1 phase are clearly sharper, indicating more phased nucleosomes. On the other hand, S phase displays the lowest SD in coverage profile, indicating lower nucleosome phasing. We have performed the same analysis for the intergenic regions and found similar results, illustrating that S phase shows in general a higher degree of chromatin disorganization (see in Discussion).

In order to check whether the cell cycle dependent chromatin condensation impacts MNase digestion levels or sequencing of certain regions, we have calculated the mean normalized coverage for each cell cycle stage. Even though it is the same along cell cycle, the boxplots of S and M phases span a wider range (**Figure 52C**), indicating that in those phases, nucleosome coverage is less uniformly distributed along yeast genes.

Overall, these observations demonstrate that chromatin organization varies along cell cycle. G1 and G2 phases have a more organized chromatin, while S and M phases are characterized by fuzzier nucleosomes along the genome and lower nucleosome occupancy in promoter regions.

4.3.3 Nucleosome Profile Transitions between Cell Cycle Stages

In order to tract cell-cycle dependent fluctuations in nucleosome profiles, we analyzed the changes between subsequent phases for individual genes. For that reason, we classify the genes according to the transitions of -1, + 1 nucleosome positioning or NFR width (**Figure 53**). Mainly, the genes conserve the nucleosome phasing and NFR states. In around 55% of the genes, -1 nucleosomes are well-positioned and 80% of the genes have +1 well-positioned nucleosomes, while 47% have both -1 and +1 well-positioned nucleosomes. Despite the high consistency among nucleosome profiles of individual genes, there is notable amount of genes showing variations. The major fluctuations are seen as a -1 and +1 nucleosome positioning change from W to F either in S and M phases, in agreement with the general variations in nucleosome maps (**Figure 51**). 141 genes change -1 nucleosome phasing from W to F at M phase and 147 genes change +1 nucleosome phasing. Although GO and pathway annotation analyzes could not find any significant enrichment in those genes, 20 of them with fuzzy -1 nucleosome have a role at M phase. Moreover, -1 nucleosome of 113 genes is well-positioned at G1 phase, becomes fuzzier at S and M phases and recovers

high phasing at G2 stage. The same behaviour is also observed for the +1 nucleosome of 71 genes. Other notable nucleosome fluctuations are the ones that have less phasing in -1 or +1 nucleosomes at G1 stage. Those genes are found to be involved in response to pheromone, conjugation and sexual reproduction, as shown in GO analysis (Appendix 4). Interestingly, the genes with fuzzy +1 nucleosomes are usually the ones activated by alpha factor pheromone. However, some genes with less phasing in -1 nucleosomes play a role in cell cycle independently of the alpha factor pathway.

	WWWW	FFFF	MMMM	WWWF	WFWW	WFWF	FFWF	WFFF	MMFM	FWWW
-1 Nuc	1789	167	141	141	137	113	76	69	62	42
+1 Nuc	2616	64	-	147	109	71	-	-	-	62

	cccc	oooo	cocc	ccco	oooc	occc
NFR	1271	1059	59	53	46	45

Figure 53: Transitions of -1 and +1 nucleosome phasing and NFR width between subsequent phases. Nucleosome profile fluctuation around TSSs is shown along cell cycle. F stands for fuzzy, W stands for well-positioning for -1 and +1 nucleosomes, while c represents closed NFR state and o is open NFR state.

In around 70% of the genes, NFR state remains the same in all cell cycle stages, among which 39% are 'open' and 32% are 'closed'. The class having closed NFR state along cell cycle, is highly enriched in TATA-box containing genes ($p<0.012$), whereas the genes with open NFR state are not among TATA-box containing genes ($p<4.94e-10$), in accordance with previous reports (Zaugg & Luscombe 2011). Moreover, the promoter regions of the genes displaying open NFR state have a high tendency to be bound by chromatin remodeler INO80 and ISW1 ($p<0.0014$ and $p< 0.003$). Interestingly, the genes changing NFR state from open to closed or *vice versa* are rare, showing that the alterations in nucleosome phasing do not alter the relative positioning distance between them.

As a next, in order to obtain a quantitative estimation of the similarity between two regions in different cell cycle stages, we defined a measurement parameter called *Coverage Difference per Base* (CDB) as:

$$\sum \frac{\left|\mathrm{cov}_x - \mathrm{cov}_y\right|}{\#\,bp}$$

where *cov* is the mean normalized coverage in the defined window for the two reference phases and #bp is the number of base pairs (bp) in the window. We have calculated and compared CDB of subsequent phases in gene body and intergenic regions (**Figure 54**). The largest cell cycle-dependent changes in nucleosome occupancy occur in G1 to S and G2 to M transitions, in both gene body and intergenic regions. On the other hand, transition between S and G2 stages does not display dramatic differences in the coverage profiles. We have analyzed genes showing the highest coverage differences between subsequent phases in more detail and observed that the top 5% of genes with highest CDB between G1 and S phases have a function in pheromone response, conjugation and sexual reproduction, as shown in GO analysis (Appendix 5). Those genes are similar to but not necessarily the same as genes with FWWW -1 and +1 nucleosome transitions (**Figure 53**). In contrast, the top 5% of genes with largest differences between G2 and M phases are highly enriched in translational elongation and translation.

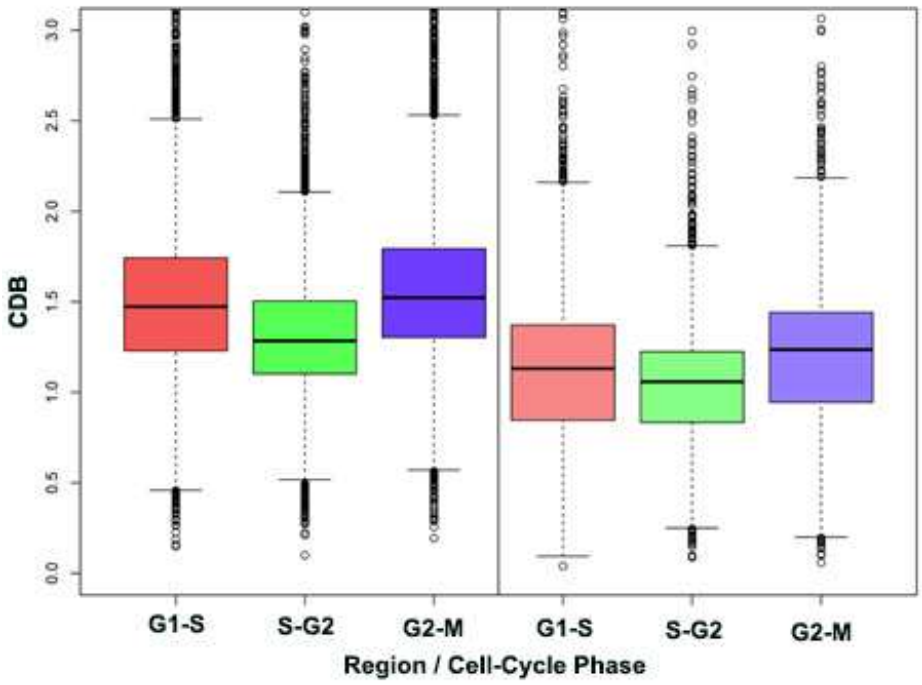


Figure 54: Coverage Difference per Base (bottom) between subsequent phases in gene bodies (left) and intergenic regions (right). Genes where one of the pairs had a mean coverage < 1 rpm were excluded. CDB of G1-S transition is shown in red, S-G2 transition in green and G2-M transition in purple.

Taken together, these results suggest that even though nucleosome profiles are quite robust along cell cycle, the nucleosome fluctuations around TSSs are mostly seen at S and M phases, where nucleosomes become fuzzier. On the other hand, the largest genome-wide variation is observed between G2 and M phases. Moreover, NFR width comparison reveals that chromatin dynamics along cell cycle impacts nucleosome phasing and occupancy but not nucleosome location.

4.4 Interplay between Nucleosome Architecture and Gene Expression

Since nucleosomes pose an obstacle for the accessibility of RNA polymerase to DNA, transcription regulation necessarily involves rearrangements of chromatin structure. To analyze the interplay between transcription and nucleosome positioning, we also extracted the RNA content of the synchronized samples in parallel with the nucleosomal DNA preparations. RNA samples were then subjected to microarray hybridization experiments to analyze gene expression data along the cell cycle. We observed that the gene expression levels generally remain constant along the cell cycle (Appendix 3) and are highly correlated between stages and between replicas as well (correlation coefficients are between 0.97 and 0.99 between stages and around 0.8 between replicas).

In order to see if there is a direct relationship between gene expression levels and nucleosome profiles around TSSs, we have checked correlation coefficients between gene expression levels and -1 or +1 nucleosome scores calculated by nucleR based on the peak shape and NFR width for all genes. However, we could not find any significant correlation between nucleosome architecture and transcription level. Therefore, as there is no general trend among genes, we turned our focus toward those particular genes that are highly and lowly expressed. We have chosen top/bottom 100 genes that are highly/lowly expressed at all cell cycle phases and compared their nucleosome profile around TSSs (**Figure 55**). In general, nucleosome occupancy around TSSs is lower in highly expressed genes. This is clearer in -1 nucleosomes, where the nucleosome occupancy in highly expressed genes is almost 50% lower than lowly expressed genes. Interestingly, despite lower occupancy, upstream nucleosomes in highly expressed genes are phased, whereas lowly expressed genes have more diffused signals, i.e. -1 and -2

nucleosome signals are merged. Moreover, NFR becomes deeper while expression decreases. There is not a clear variation between cell cycle stages in none of the cases, implying that nucleosome architecture around TSSs might determine the expression levels or *vice versa* quite independently of the cell cycle phase.

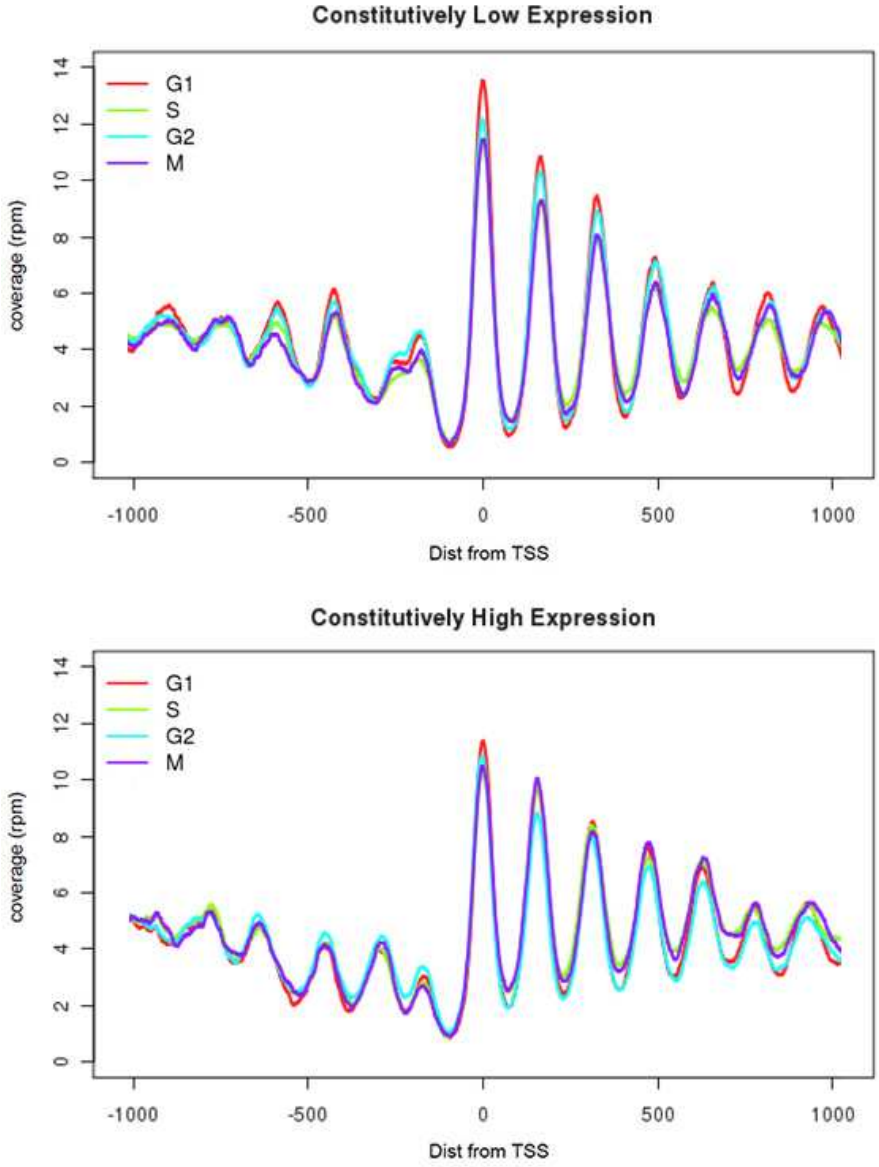


Figure 55: Nucleosome profiles of lowly and highly expressed genes. Coverage maps per base pair were calculated and averaged across top/bottom 100 highly/lowly expressed genes around TSSs at G1 (red), S (green), G2 (blue) and M (purple) phases.

In an attempt to study the relationship between nucleosome occupancy in coding regions and expression levels, we compared the gene body mean nucleosome coverage of highly and lowly expressed genes (**Figure 56**). We noted that highly expressed genes display higher coverage, hence they are more occupied by nucleosomes, similar to what was reported before (Lee et al. 2007). We observed the same behavior at each cell cycle stage. Surprisingly, while coverage of lowly expressed genes at each stage is quite similar to each other, highly expressed genes display much larger coverage at S and M phases.

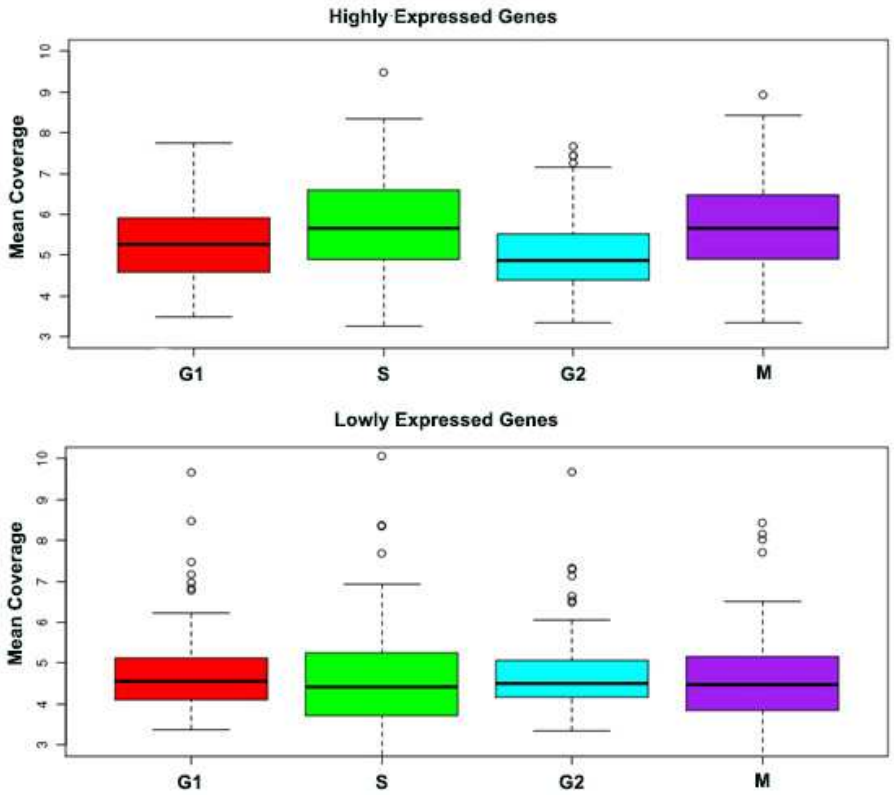


Figure 56: Relationship between Coverage Difference per Base (CDB) and expression levels. Boxplots show the coverage distribution in highly (upper) and lowly (lower) expressed genes at each cell cycle stage.

In order to gain a deeper insight into the interconnection between nucleosome architecture and expression plasticity, we analyzed the variations in gene expression between subsequent phases (Figure 57). The largest differences in gene expression are observed between G1-S phases, followed by G2-M phases, where genome-wide CDB variations are also high.

On the other hand, the expression variation is very low between S and G2 phases, where the CDB variation is the lowest (Figure 56). This observation indicates that some the nucleosome coverage changes might be explained by gene expression variations along cell cycle. Moreover, we separately analyzed 10% of the genes showing the highest CDB SD and those with highest expression plasticity along cell cycle. Interestingly, the genes with nucleosome profile variations have a high tendency to change their expression along cell cycle ($p < 5.6e-08$, Appendix 6). Moreover, GO analysis shows that the genes displaying high CDB and expression change are enriched in cell cycle-dependent activities (Appendix 7), further confirming the link between transcription and chromatin.

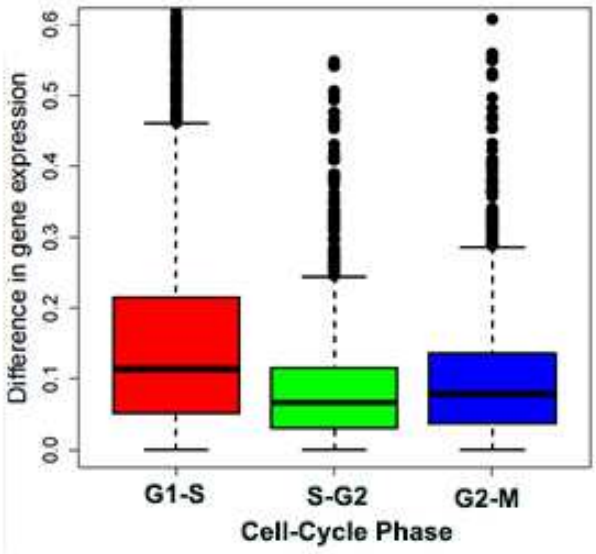


Figure 57: Gene expression differences between subsequent phases. G1-S transition is shown in red, S-G2 transition in green and G2-M transition in blue.

Overall, our results show that even though there is not a global relation between nucleosome organization and gene expression, highly expressed genes display different nucleosome organization comparing to lowly expressed ones. Moreover, the genes displaying the largest nucleosome profile changes are also showing high expression plasticity and are usually cell-cycle-regulated genes.

4.4.1 Cell Cycle-Regulated Genes

We have extended our study on interplay between transcription and nucleosome profile by examining individual genes based on their nucleosome architecture and expression status. Particularly, we examined the genes showing both differential nucleosome architecture and gene expression throughout cell cycle.

The majority of the genes involved in the alpha-factor pheromone mating are active in G1-arrest which is coupled with a significant change in the nucleosome profile at this stage. The interplay between the gene expression and chromatin structure is particularly evident here, because the achieved synchrony is higher at that stage and those genes are induced by an external

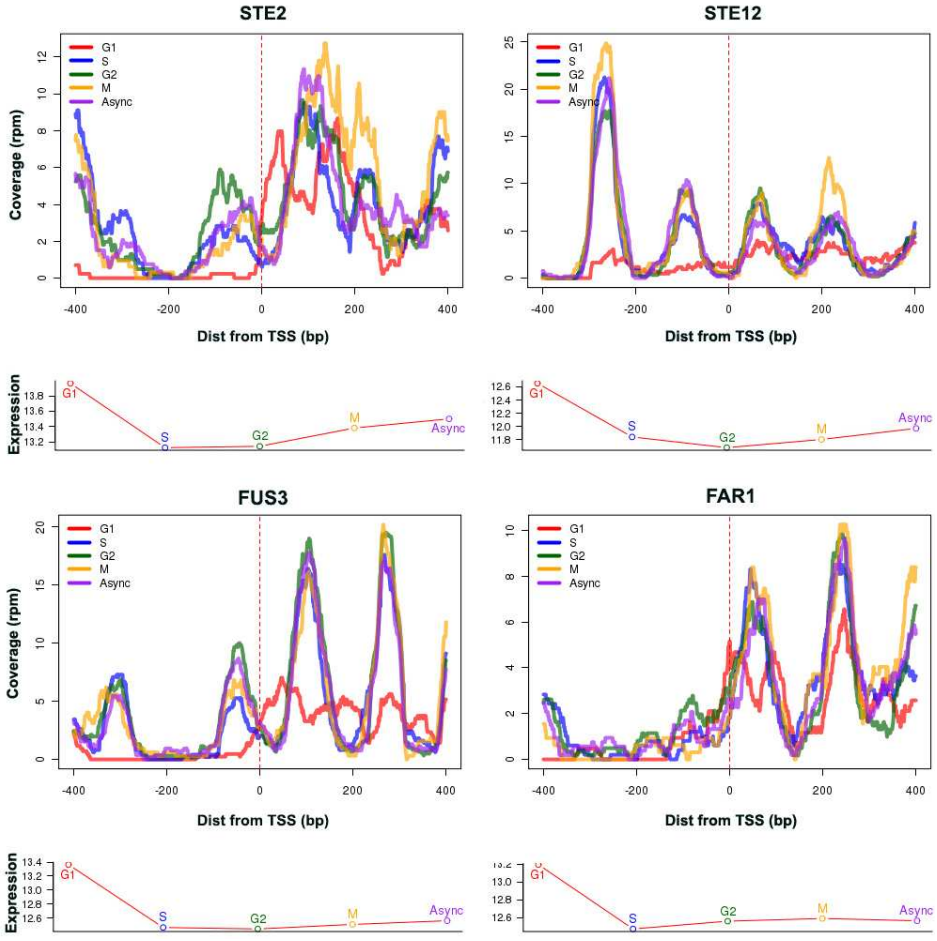


Figure 58: Nucleosome architecture and expression plasticity of genes involved in alpha-factor pheromone mating and its desensitization response. Gene expression levels are

indicated as log2 values of the hybridization ratios from the Affymetrix GeneChip Yeast Genome 2.0 arrays.

stimulus. First, alpha-factor pheromone is captured by the receptor Ste2, and activates Ste2. The activation of Ste2 in G1 arrest is coupled with an almost complete eviction of nucleosomes in the TSS upstream region and a significant increase in the fuzziness of the +1 nucleosome (**Figure 58**). Ste2, in turn, activates heterotrimeric G-protein, and eventually the upstream component of MAPK cascade. The activation of Fus3 (MAPK) consecutively phosphorylates and stimulates Ste12 (**Figure 58**) and Far1 to promote cell cycle arrest at G1 phase. The activation of Fus3, Far1 and Ste12 corresponds to very lowly occupied or completely evicted -1 and +1 nucleosomes.

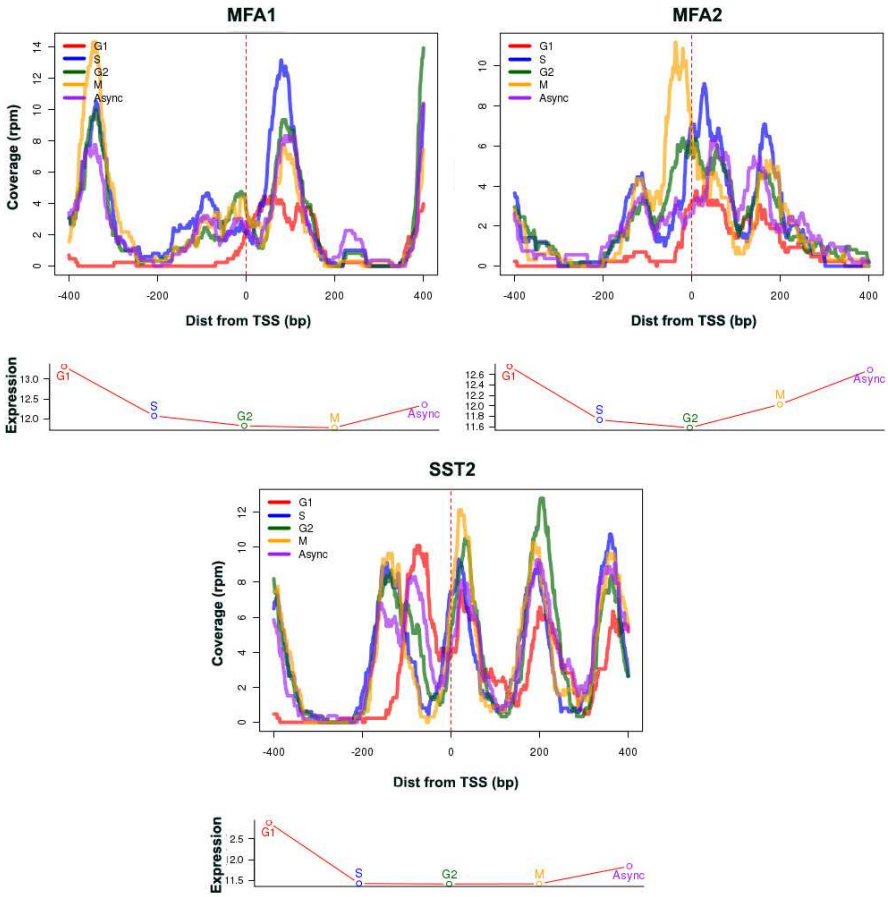


Figure 59: Nucleosome architecture and expression plasticity of genes involved in alpha-factor pheromone desensitization response.

Furthermore, yeast cells usually recover from mating pheromone alpha-factor-induced division arrest by desensitization. Hence, genes encoding the alpha-factor (MFA2 and MFA1) or those that regulate desensitization to alpha-factor such as SST2 (GTPase-activating protein) are highly transcribed in the presence of alpha-factor in G1 arrest (**Figure 59**). Once more, increase in the expression of these genes is coupled with a dramatic reduction in the -1 and +1 peaks, shift in the positioning or even a total eviction of these nucleosomes, providing further evidence on the key role of chromatin plasticity in this signal cascade.

Besides the genes involved in alpha-arrest mechanism, the genes having a role in cell cycle are composing the other major part of this class. Once the cells are released from the alpha-factor, Cln1 and Cln2 cylins activate Cdk1 to complete the progression through the START by blocking the Sic1 inhibitor that suppresses S phase Cdk1 activity. Hence, Sic1 has lower expression at G1 and S phase, whereas the maximum expression is at M phase, where the -1, +1 and +2 nucleosomes have lower occupancy (**Figure 60**). The destruction of Sic1 allows the transcription of Clb5 and Clb6. They are active at the beginning of S phase and their expression drops right after, and increases again at M phase. This expression pattern is seen in our replicas as high M phase expression, since the samples are collected toward the mid-S phase. As seen in Clb6 (**Figure 60**), the occupancies of -1 and +1 nucleosomes are lower at M phase corresponding to their expression status. Clb5 and Clb6 directly promote initiation of DNA replication by triggering pre-replicative complex in replication origins. Once the complex is active, helicase maintenance (MCM) proteins unwind DNA to expose template DNA. MCM is activated at M/G1 phase. MCM3 and MCM7 subunits are good examples of the high interplay between expression and the nucleosome profile. There is a shift away from TSS in the downstream nucleosomes at G1 and M phases (**Figure 60**).

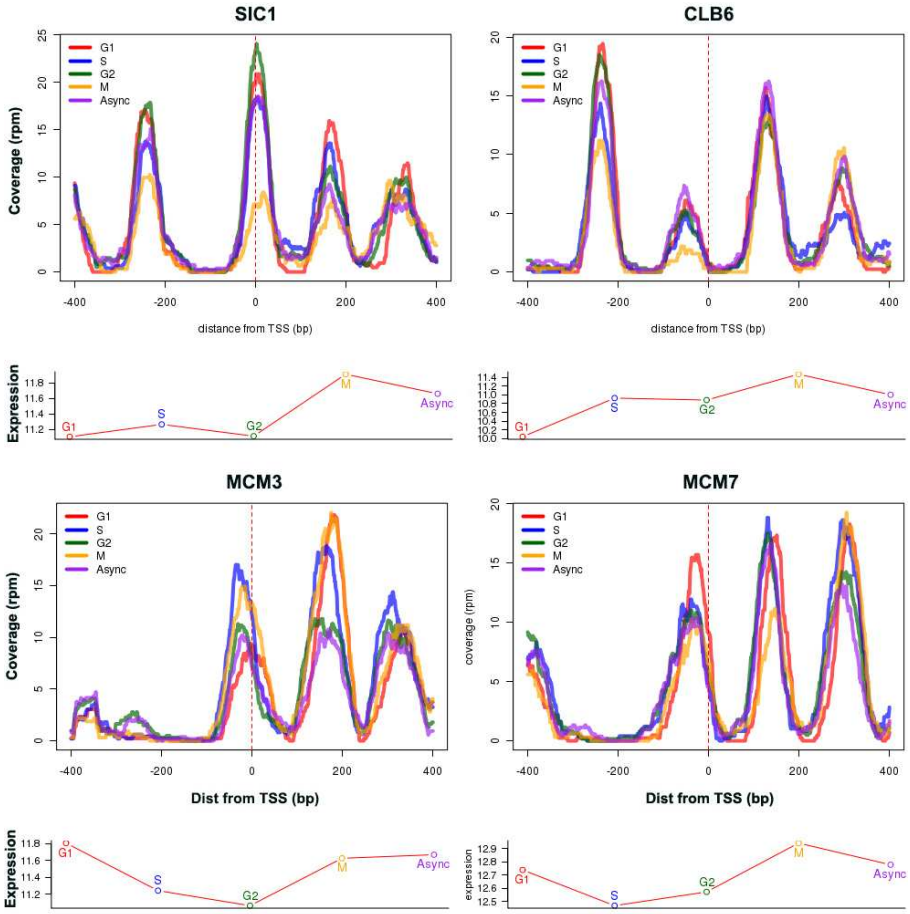


Figure 60: Nucleosome architecture and expression plasticity of SIC1, CLB6, MCM3, and MCM7.

Once the replication is completed, the cell cycle progression is under control of other B type cyclins. Clb1 is one of these B type cyclins which fall into a class whose expression is activated from late S phase to M phase. Apart from CLB1, several other known genes exhibit a similar temporal expression pattern: SWI5, ACE2, and CDC5. Interestingly, it was shown that the transcriptions of these genes are directly controlled by a general transcription factor MCM1 (Althoefer et al. 1995). Not surprisingly, their nucleosome profiles show a similar pattern variation. The nucleosome profiles at S, G2 and M phases, for which the expression is higher, show a different phasing than G1 phase with lower expression (**Figure 61**). That differential nucleosome profile is clearer for Clb1, where the nucleosome

profile at G1 phase is distinct when compared with the nucleosome profiles at S, G2 and M phases.

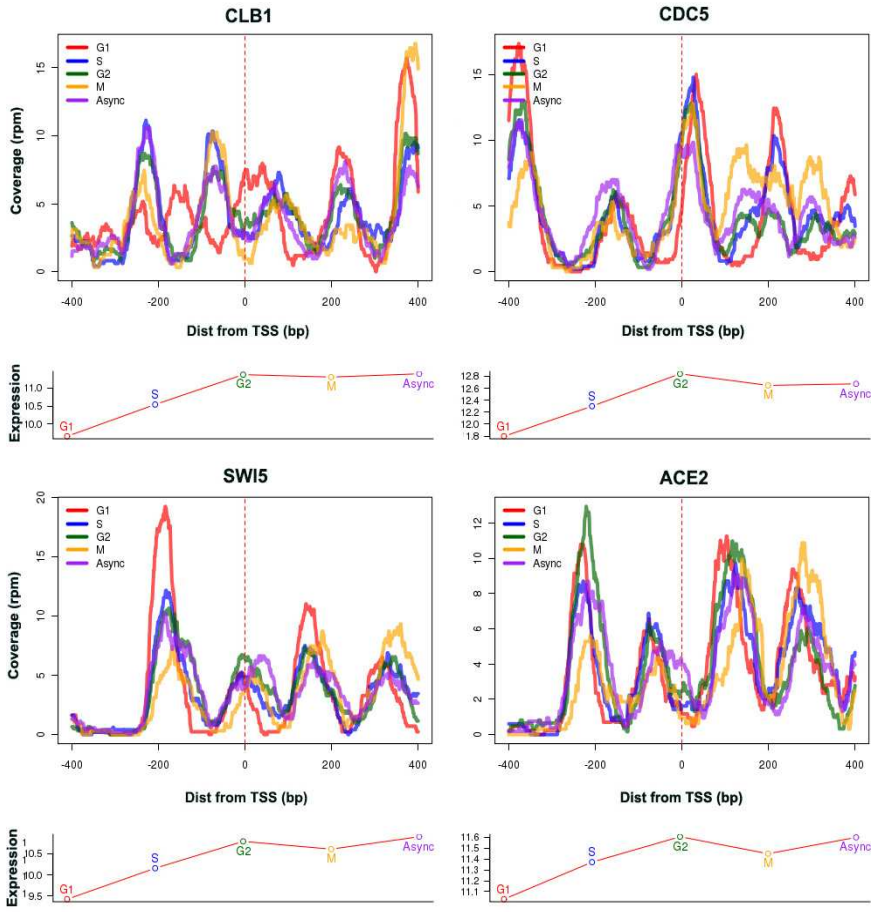


Figure 61: Nucleosome architecture and expression plasticity of CLB1, CDC5, SWI5, and ACE2.

5 Nucleosome Architecture around Replication Origins

5.1 Replication Origin Nucleosome Profile Along Cell cycle

The nucleosome pattern around yeast origins have been studied in great details (Berbenetz et al. 2010; Eaton et al. 2010; Hoggard et al. 2013). Even though the replicatory protein binding and their activation is a dynamic process, limited work has been done on chromatin dynamics around replication origins along cell cycle and most of the reported origin nucleosome patterns are from asynchronous or G1 arrested cells. For that reason, we have examined and compared nucleosome positioning maps around replication origins at every cell cycle stage. The origin nucleosome profiles were generated from synchronized cells in triplicates, described in Materials&Methods, 2.2&2.5, by aligning 253 annotated origins relative to ACSes, where ORC binds (Eaton et al. 2010). Similar to TSSs profiles, we refer to the first upstream nucleosome as -1 nucleosome and the first downstream nucleosome as +1 nucleosome. Expectedly, the nucleosome pattern of replication origins is quite similar to that of TSSs in all replicates along cell cycle (**Figure 62**). However, NFRs at ACS, unlike TSS profiles, are flanked by well-positioned nucleosomes in a symmetrical manner, and NFRs around origins are wider than 5' NFRs. Furthermore, even though the general profile is quite similar along cell cycle, there are some slight variations at distinct cell cycle stages. The nucleosome occupancy is lower at G1 and S phases compared to G2 and M phases. Replicate 1 and 3 exhibit the lowest occupancy at G1 phase, when pre-RC is formed and the highest occupancy at G2 phase, when all origins are fired and the replication complex is disassociated. However, replicate 2 displays the lowest nucleosome occupancy around replication origins at S phase, followed by G1 phase. Interestingly, S phase of the same data set illustrates the lowest nucleosome occupancy around TSSs and highest MNase sensitivity based on deep sequencing and MNase digestion assays (**Figure 45&Figure 46**). Therefore, the low nucleosome occupancy of replicate 2 at S phase around origins might account for the genome-wide effect of chromatin structure alterations. Furthermore, NFR around replication origins is marginally wider at G1 and S phases than G2 and M phases in all replicates. Even though the difference is insignificant, the hindering effect of profile averaging on individual nucleosome variation should not be neglected.

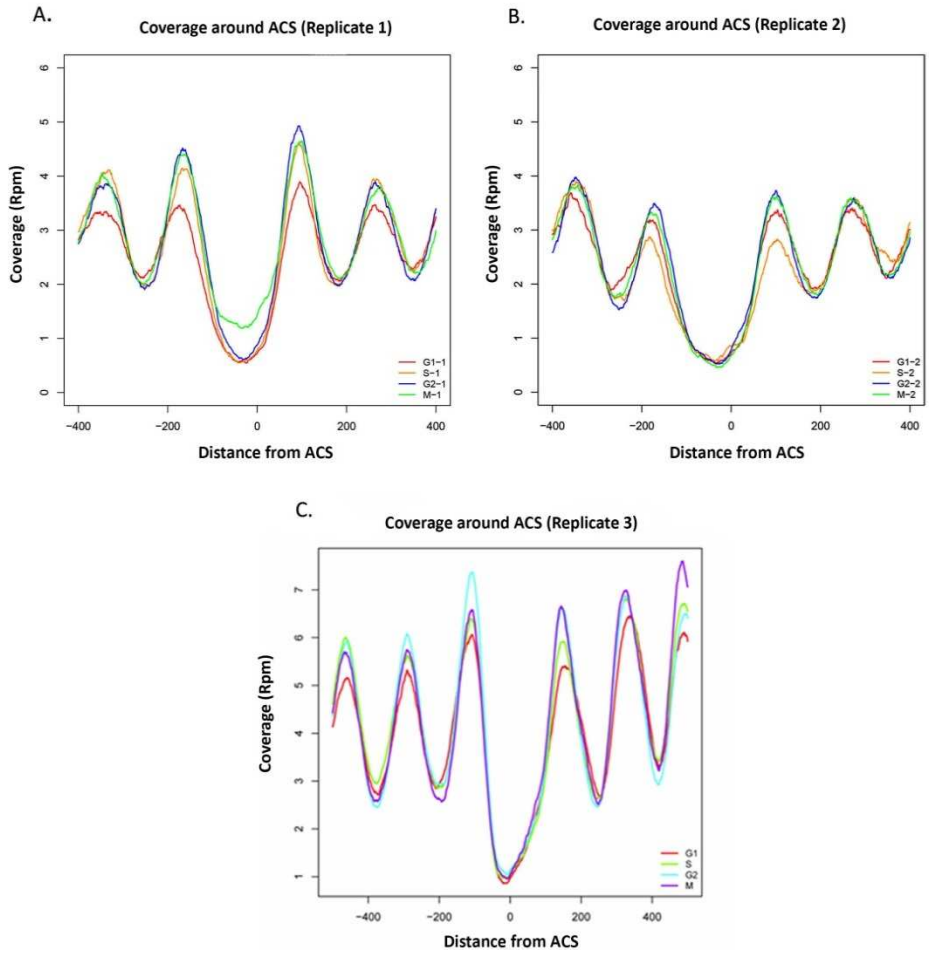


Figure 62: Average nucleosome profiles around replication origins relative to ACS along cell cycle of replicates 1(A), 2 (B), 3(C). Coverage maps per base pair were calculated and averaged across all yeast genome around ACSes.

To minimize the profile averaging effect and to examine origin profiles individually, we have generated origin heatmaps relative to ACS that are ordered by increasing NFR width (**Figure 63**). Despite the high similarity among the heatmaps of distinct stages, the slight differences are notable among origin profiles. The majority of the origins at G1 stage display high nucleosome phasing at NFR flanking regions and their NFRs are completely depleted of nucleosomes. On the other hand, nucleosome phasing decreases at S, G2 and M phases and NFRs are slightly occupied by nucleosomes in some origins, albeit not strongly.

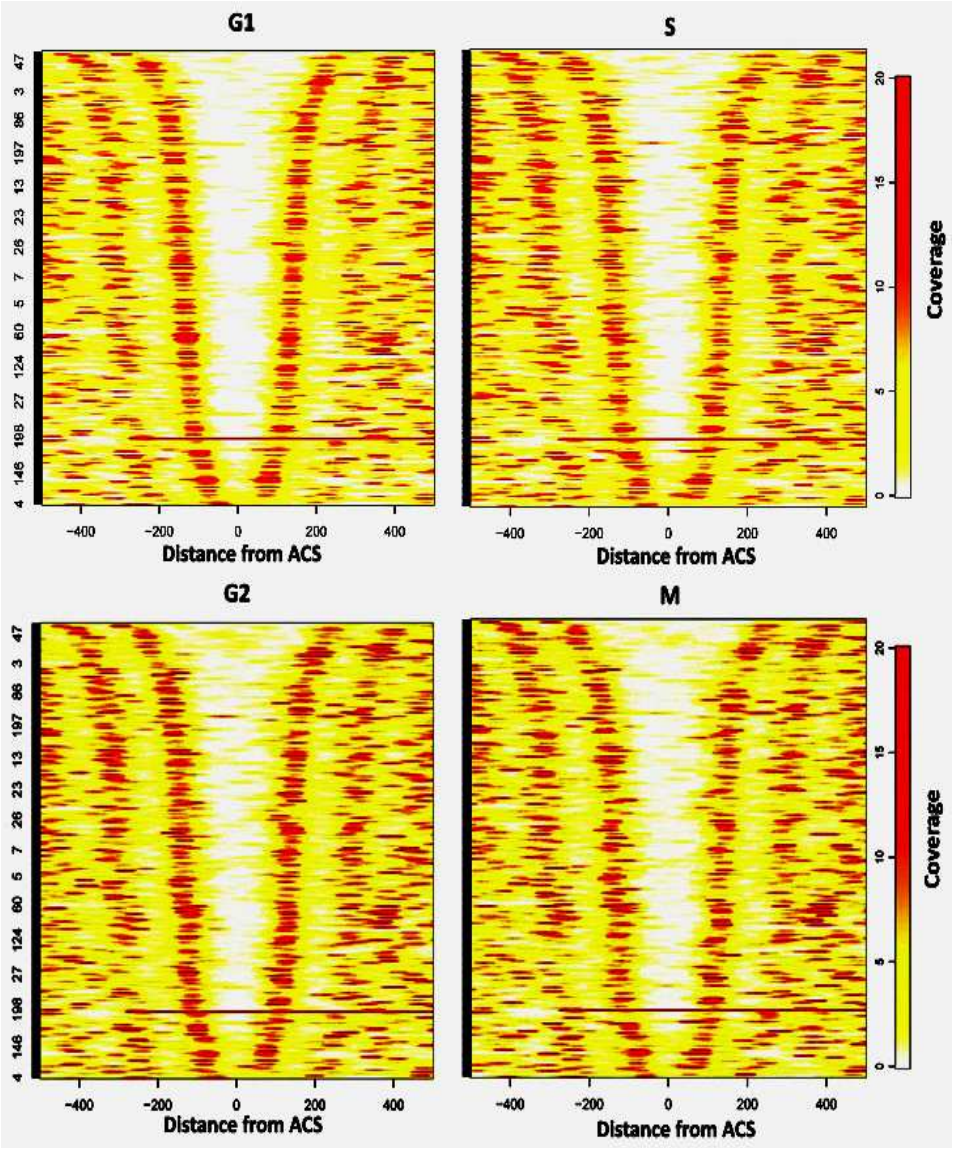


Figure 63: Heatmap of nucleosome occupancy around replication origins along cell cycle. 243 annotated replication origins are aligned on Y axis and flanked from smallest to largest NFR width.

To further investigate the NFR occupancy differences, we have compared the NFR widths along cell cycle (**Figure 64**). We observed that NFR width follows a similar distribution with a peak around 300 bp along cell cycle (Appendix 8). However, NFRs tend to be larger in G1 phase compared to other stages. S phase displays very similar NFR distribution to G1 phase, while the lowest NFR widths occur at G2 phase. Even though the maximum median difference

along cell cycle is only around 20 bp, it is highly significant ($p < 2x10^{-9}$). Overall, heatmaps and NFR median comparison demonstrate that when pre-RC is formed at G1 phase, ACS is completely depleted of nucleosomes and once the complex disassociates, nucleosomes slide and marginally occupy the ACS region. Yet, the occupation is still not high since ORC remains bound to ACS region.

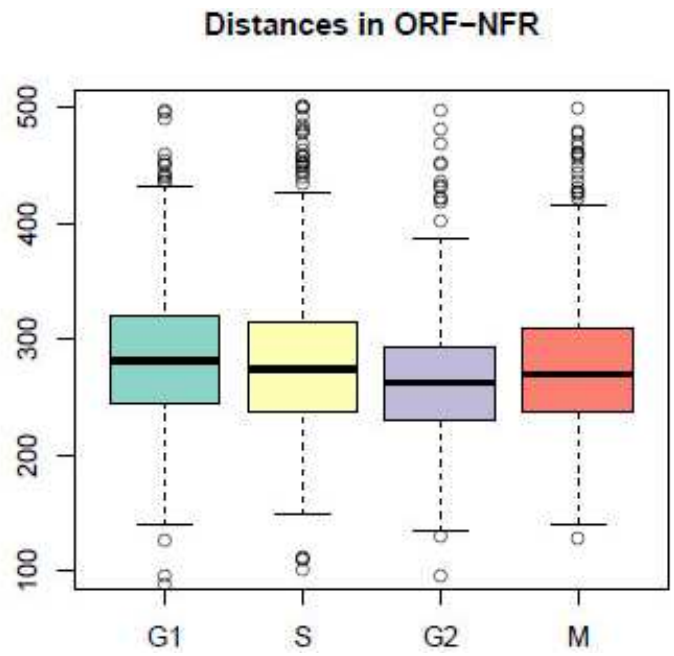


Figure 64: NFR width median at ACS along cell cycle. NFR width median is shown at G1 (red), S (purple), G2 (green) and M (cyan) phases.

Besides the slight variations along cell cycle, the nucleosome pattern diversity among individual origins is remarkable based on the origin heatmaps, as reported before (Berbenetz et al. 2010). Explicitly, some origins display weaker or stronger nucleosome positioning, with either fuzzier nucleosomes or wider or narrower NFRs that rank from 80 bp to 600 bp. In order to analyze the variations in the patterns in more detail, we have plotted the nucleosome profiles of 4 representative origins, ARS416, ARS305, ARS1323 and ARS1120. As shown in **Figure 65**, while ARS416 and ARS305 have wider NFRs of about 200 bp, ARS 1323 and ARS1120 show much narrower NFRs. Moreover, the diversity is notable in the flanking nucleosomes: with high occupancy like in ARS416 or with relatively lower occupancy like in ARS305 and ARS1120. Interestingly, the differential

nucleosome organization at G1 phase was also remarkable in the individual patterns, ARS 305, ARS1323 and ARS1120.

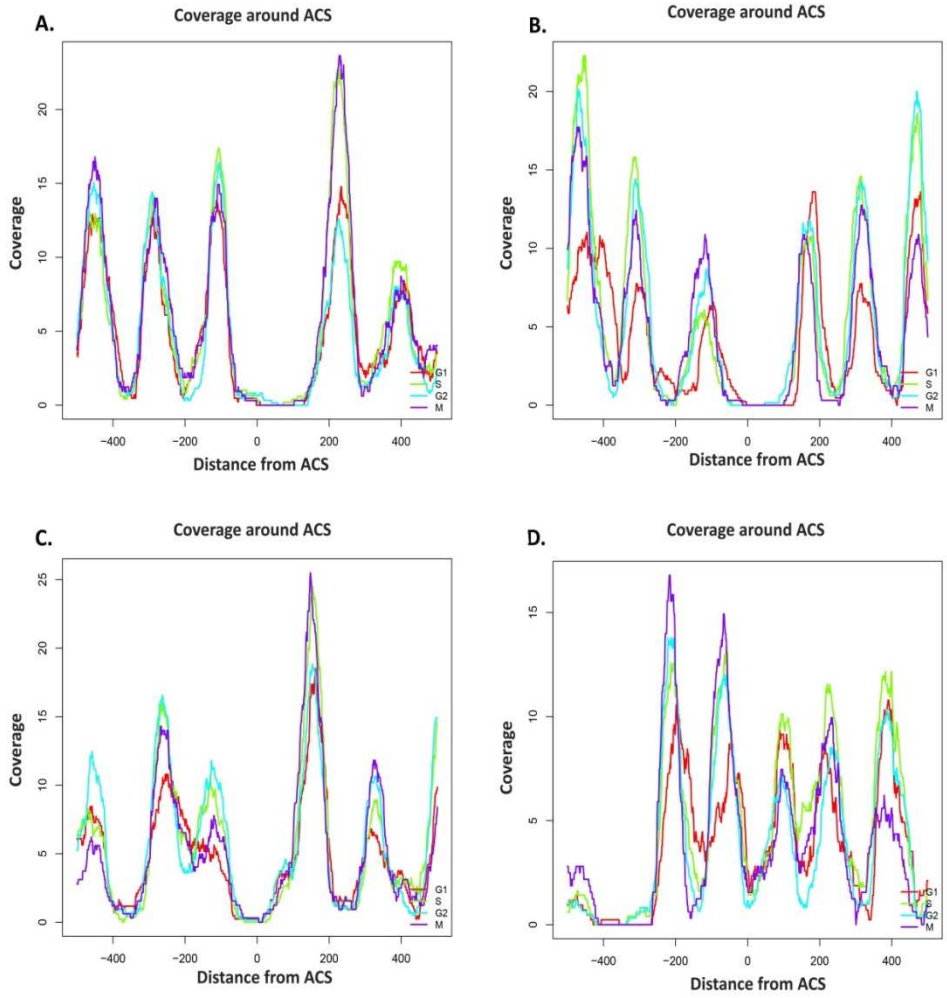


Figure 65: Nucleosome organization around individual origins. ARS416 (A), ARS305 (B), ARS1323 (C) and ARS1120 (D) origin nucleosome profiles are shown relative to ACS.

5.2 Replication Origin Timing In Relation to Nucleosome Organization

To check if the nucleosome pattern variation is related to origin timing, we have first examined the relationship between NFR width and origin firing. The replication timing data is taken from Yabuki *et al.* (Yabuki et al. 2002) which provides 256 origin timings ranking from 16.8' to 36.9' based on copy number variation, from one to two copies during DNA replication. We have

examined 168 origins, since we did not obtain reliable nucleosome coverage for the remaining 88 origins. For simplicity, only the relation between origin timing and NFR width distribution at G1 phase is shown, since the other phases provided very similar results. Intriguingly, NFR width distribution does not seem to be related to origin timing **(Figure 66A**). As a next step, we have directly compared NFR width of early and late origins to check whether they display a random pattern. For that reason, we have classified the origins as early and late by selecting the earliest and latest 10% of the total 168 origins. **Figure 66B** demonstrates the median comparison of early and late origins at each cell cycle stage. As observed in the plots, the NFR width median does not change among early and late replication origins in none of the cell cycle stages.

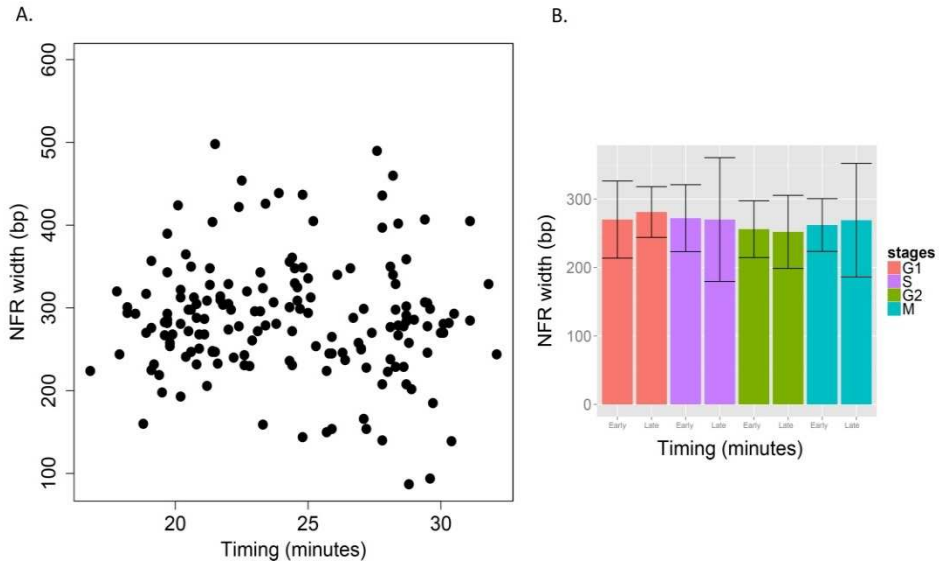


Figure 66: Correlation between origin NFR width and replication timing. A. The plot shows the relation between firing timing and NFR width distribution at G1 phase. B. The median NFR width at early and late replication origins are compared along cell cycle.

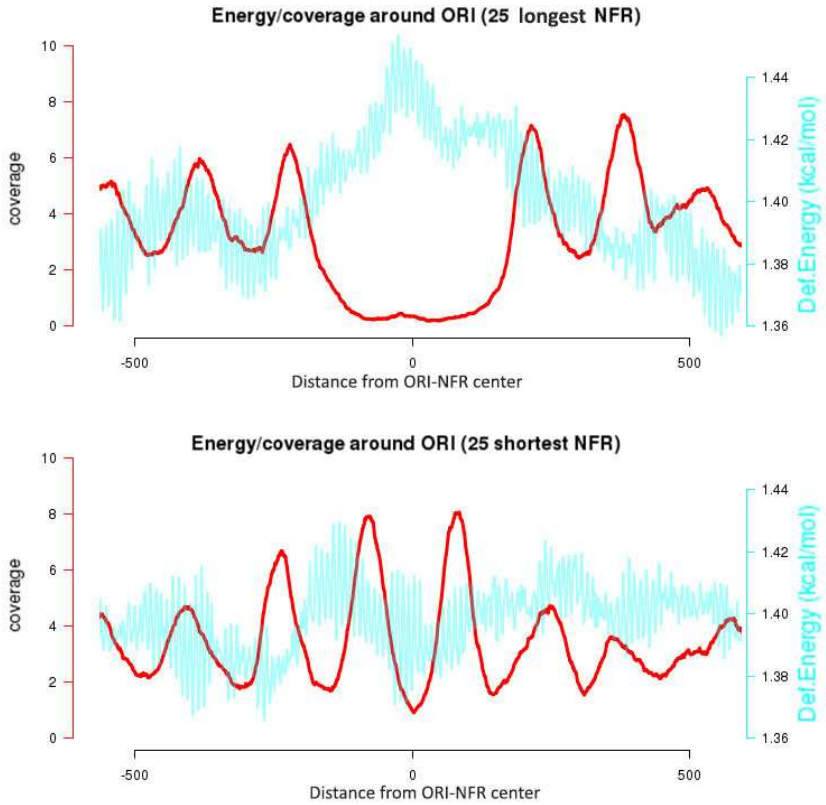


Figure 67: Deformation energy and coverage profiles around replication origins. Total deformation energy was calculated and averaged across genome around top 10% (top) and bottom 10%(bottom) replication origins in terms of NFR width.

In an attempt to examine whether DNA physical properties play a role in determination of NFR width around origins, we plotted the average deformation energy around NFR of origins and compared it with nucleosome profiles around top and bottom 10% origins in terms of NFR width (**Figure 67**). We observed that origins with wider NFRs have a higher deformation energy values, implying that these regions are more difficult to wrap around a histone core. On the other hand, origins with narrow NFRs display low deformation energy, enabling nucleosome formation. This observation indicates NFR width distribution could be determined by DNA intrinsic properties, independent of origin activities.

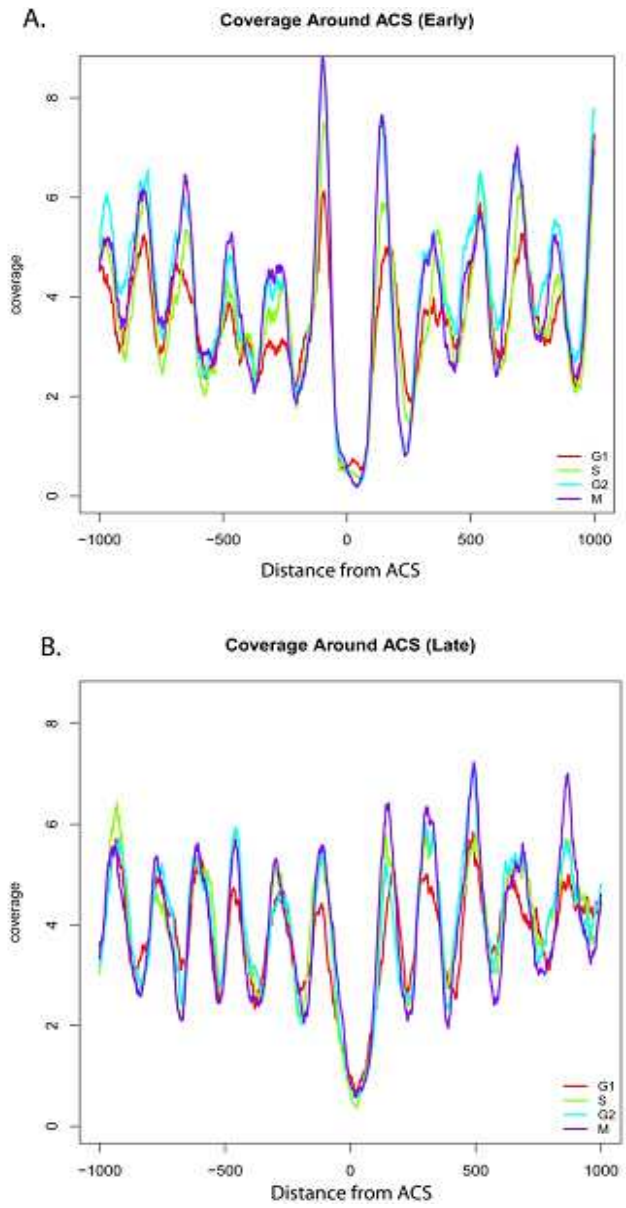


Figure 68: Nucleosome profiles of early (A) and late (B) replication origins relative to ACSes. Early and late origins are defined by selecting top 10% (A) and bottom 10% (B) origin timing annotated by Yabuki et al., 2002. Their nucleosome occupancy profiles were plotted relative to ACSes.

To examine if the origin timing influences nucleosome positioning and occupancy, we have also compared nucleosome profiles of early and late origins. The average early origin profiles show higher nucleosome occupancy, most clearly evident at -1 and +1 nucleosomes (**Figure 68**). Moreover, NFRs

of early origins are deeper and located asymmetrically toward downstream and their flanking nucleosomes have higher phasing, characterized by sharper peaks and more well-defined linker regions. On the other hand, late origins display symmetrically positioned NFRs, flanked by less well-defined nucleosomes. Furthermore, early origins display lower nucleosome occupancy around ACSes at G1 and S phases compared to the other phases, whereas late origins show striking low occupancy only at G1 phase. In order to minimize the profile averaging effect, we compared the fuzziness of eary and late origins at G1 phase for the simplicity. Boxplots in **Figure 69** clearly demonstrate that early firing origins are better phased (higher nucleR scores). Moreover, the well positioning in early firing origins is quite robust along the cell cycle, as indicated with a lower SD. In summary, even though the general nucleosome pattern is conserved in both early and late firing origins, early origins tend to be mostly characterized by well positioned nucleosomes around the ACS region.

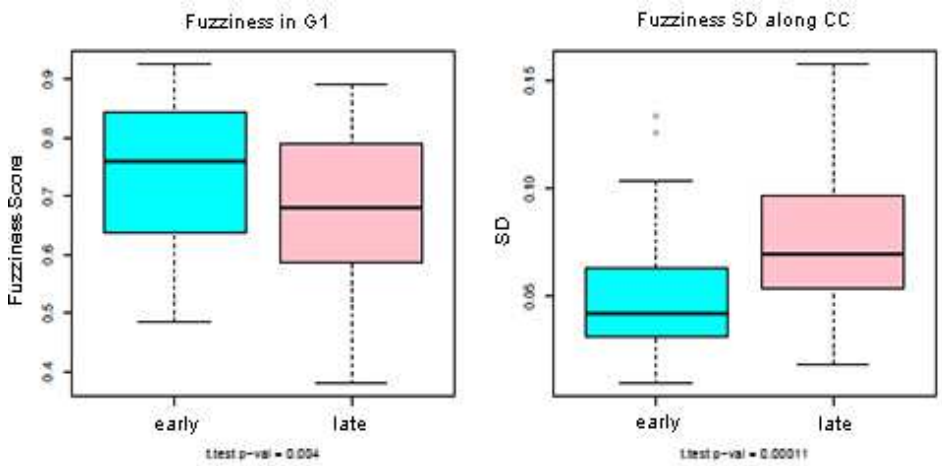


Figure 69: Fuzziness score around early and late origins. The average nucleR scores of -1 and +1 nucleosomes are calculated and the distribution among early and late origins are indicated in the boxplot (left). Variations in fuzziness score along cell cycle is shown on the right boxplot.

To check if the higher phasing of early origins is epigenetically encoded, we have compared H2A.Z histone variant occupancies in early and late origins, since H2.A.Z has a role in nucleosome phasing. Genome-wide H2.A.Z occupancy data is taken from (Albert et al. 2007) and plotted for early and late replication origins (**Figure 70**). H2A.Z occupancy at -1 and +1 nucleosomes display very similar patterns in both early and late origins,

where -1 nucleosomes have higher H2A.Z occupancy at early origins than late origins when compared to +1 nucleosomes. Moreover, we observe a similar trend in NFR location symmetry differences in H2A.Z occupancy profiles. Altogether, differences in H2A.Z occupancy might partially explain the high nucleosome phasing at early replication origins, which may perhaps play a regulatory role in early origin firing and efficiency. On the other hand, late origin firing and activity could be less nucleosome dependent.

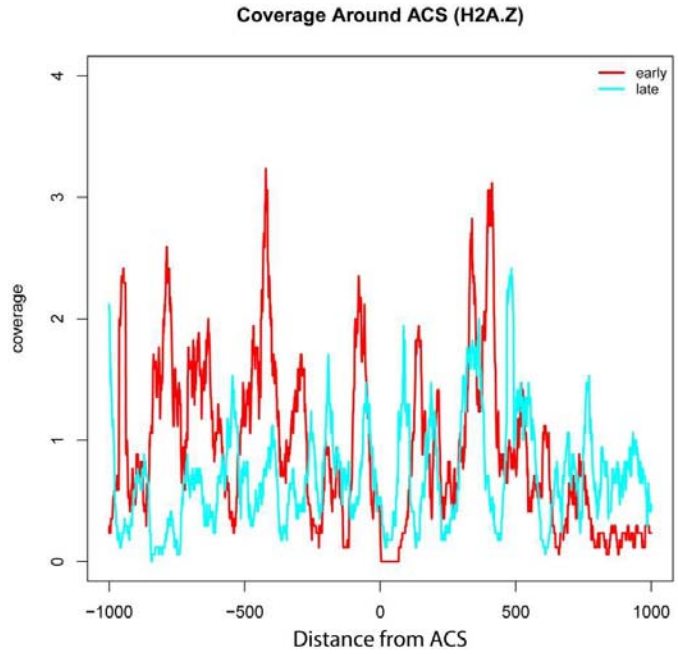


Figure 70: H2A.Z occupancy at early and late replication origins. H2A.Z occupancy (Albert et al., 2002) are shown at top 10% (red) and bottom 10% (late) timing of origins.

6 Centromeric Nucleosome

The exact composition and structure of centromeric nucleosomes in *S. cerevisiae* are still under discussion. Conflicting models have been proposed based on various techniques, most of them involving *in vitro* reconstitution assays, introduction of C-terminal tags to centromeric histone proteins or usage of fixative agents, which might bias and not reflect the actual centromeric nucleosome organization. Therefore, in order to obtain an accurate picture of centromeres, we studied the centromeric nucleosome positioning and organization under physiological conditions.

6.1 Centromeric Nucleosome Positioning

We first analyzed centromeric nucleosome positioning along cell cycle by comparing nucleosome profiles around centromeric DNA, which were generated from synchronized cells in triplicates, as described in Materials&Methods 2.2&2.5. We aligned the 16 centromere profiles relative to centromeric DNA and analyzed the positioning in three replicates (R1, R2 and R3; **Figure 71**). Notably, centromeric nucleosome is located at the centromere mid-point (point 0), consistent with its reported sequence dependence. Moreover, centromeric nucleosome peaks are shorter than canonical nucleosome peaks, as reported earlier (Dalal et al. 2007; Cole et al. 2011). There are slight differences in nucleosome occupancy along cell cycle; although there is not a consistent trend between replicates. On the other hand, one of the highest MNase-sensitive nucleosome map replicates displays an outstanding difference in centromere signal at S phase, with a nucleosome positioning around 170 bp and an occupancy of three times higher than that at G1, G2 or M phases. However, this observation could not be reproduced in the other replicates.

Since the average profiles might mask the individual centromeric nucleosome organization, as mentioned earlier, we then studied in more detail the individual centromere profiles. Indeed, we observed that some chromosomes are completely depleted of centromeric nucleosome signals, while some chromosomes have very high centromeric nucleosome occupancy, thus biasing the average profiles. Furthermore, individual profiles provide examples of ideal nucleosome positioning, implying that all of the nucleosomes are positioned exactly at the same place within a population.

Figure 72 shows two representative centromeric nucleosomes from chromosomes 8 and 13, with an ideal rectangular instead of a peak shaped signal. The invariant positioning of centromeric nucleosomes at each chromosome further confirms their sequence-dependent localization at centromeric DNA.

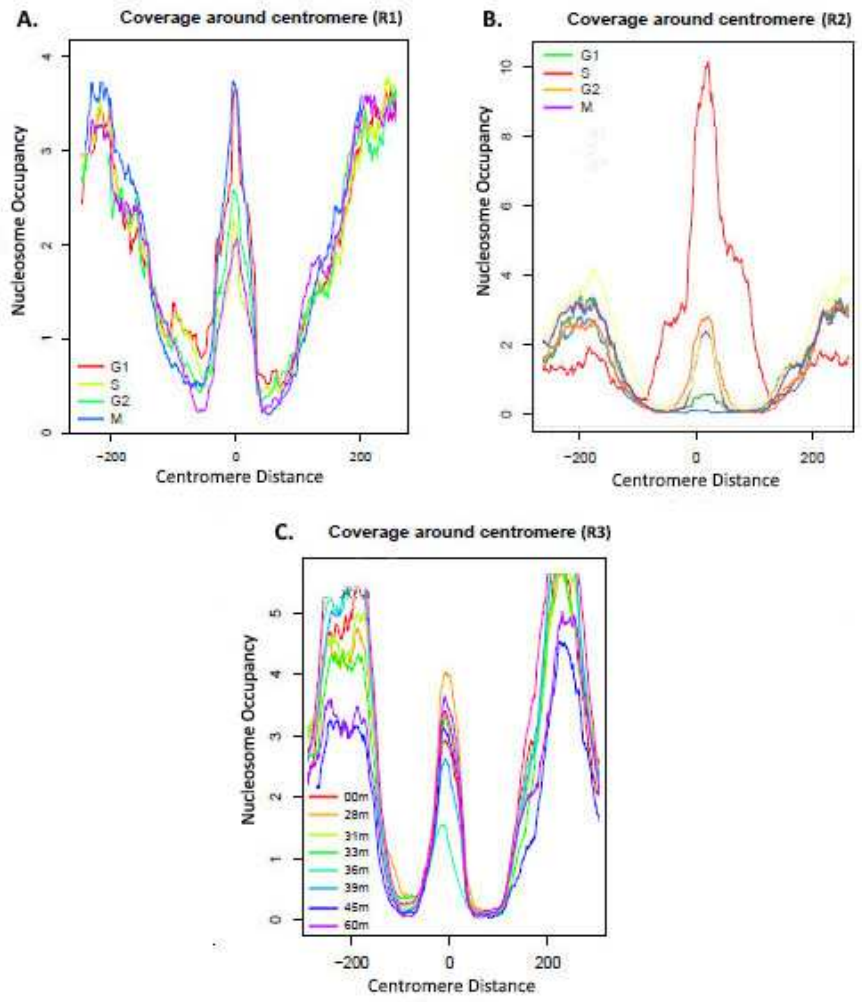


Figure 71: Average centromeric nucleosome profiles around centromeric DNA along cell cycle of replicate 1(A), 2(B), and 3(C).

Even though the centromere occupancy slightly differs on each chromosome at different cell cycle phases, we were unable to reproduce any trend in nucleosome occupancy variations along cell cycle and therefore, to obtain any conclusive results about centromeric nucleosome dynamics.

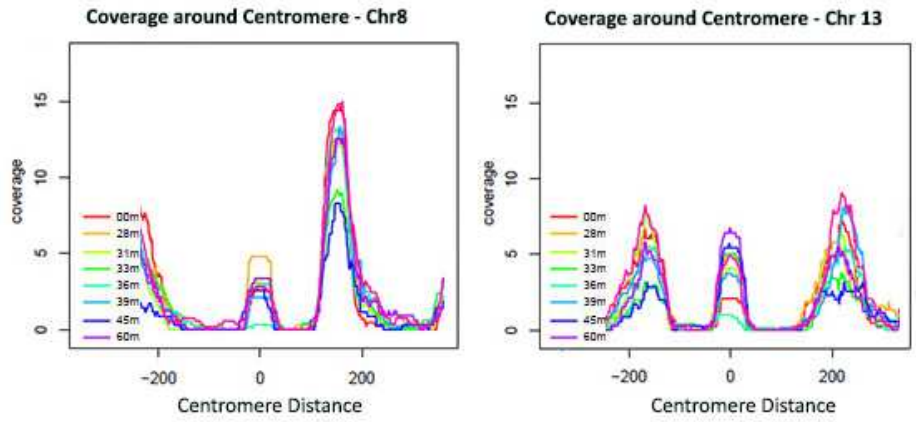


Figure 72: Nucleosome profile along cell cycle around centromeric DNA on chromosome 8&13.

6.2 Characterization of Centromeric Nucleosome

Since it is well established that centromeric nucleosomes show a higher MNase digestion sensitivity as compared to canonical nucleosomes (Cole et al. 2011), some centromeric nucleosomes might have disassociated during the experimental procedures. Furthermore, even if MNase-Seq derived maps give information about the location and the size of centromeric nucleosomes, they are not informative about nucleosome protein content and organization. Therefore, we attempted to overcome these issues using a modified nucleosome preparation protocol followed by a sucrose gradient centrifugation, with the aim to purify centromeric nucleosomes in their physiological environment (Materials&Methods 2.3). Accordingly, nuclei were prepared as described (Material&Methods 2.3.1), digested with MNase, yielding mostly mono- and di-nucleosome fragments and separated by 5-25% sucrose gradient centrifugation. Expectedly, the denser nucleosome particles like di- or tri-nucleosomes are present in the later fractions, while mono-nucleosomes or even smaller chromatin particles appear in the upper layer of the sucrose gradient (Appendix 9). On the other hand, centromeric nucleosomes appear in the fractions containing mono-/di-nucleosome fragments, verified by PCR using sequence-specific oligonucleotide primers for centromeric DNA sites at chromosomes 4 and 6 (Appendix 9). Thus, even though 5-25% sucrose gradient is able to separate mono- from di- or tri-nucleosomes, it fails to separate canonical from centromeric nucleosomes. To increase the separation resolution, we then

lowered the gradient range to 5-15% in further experiments. In this case, the separation of mono- and di-nucleosomes is poorer and they mostly concentrate in fractions 4, 5, 6 and 7 (**Figure 73**). Interestingly, standard PCR of each fraction using centromeric primers: 5´CEN4, 3´CEN4; 5´CEN6, 3´CEN6, 5´CEN14 and 3´CEN14 (which yield centromeric DNA products of around 110 bp), displays centromeric signals in fractions 4-8, where most of the chromatin is also present. However, centromeric nucleosome signals appear in fractions 9, 10, 11, 12 and 16 as well, where canonical nucleosomes are not in abundant.

As a next, we evaluated the protein contents of centromeric nucleosome enriched fractions by silver staining inspection and further analyzed by mass spectrometry (Appendix 11). Nevertheless, the scarcity of centromeric nucleosomes (only 16 per cell) and the unavailability of an anti-histone Cse4 specific antibody make the detection of centromeres very hard. Indeed, due to the poor enrichment of centromeric nucleosomes and the protein contamination by external factors, we were not able to detect the centromeric histone protein variant Cse4 by mass spectrometry. Moreover, not all of the histone proteins were able to be visualized, even in enriched mono- and di-nucleosome fractions.

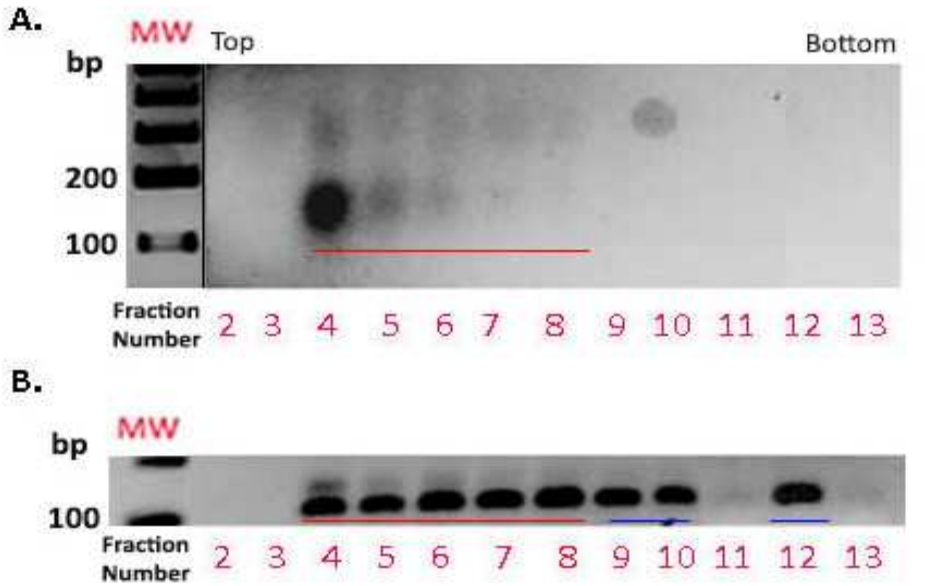


Figure 73: Sucrose gradient (5-15%) fractionation of MNase digested chromatin. A. Portions of fractions were analyzed on 2% agarose gel. B. The presence of centromeric nucleosomes was determined by standard PCR using 5´CEN4 and 3´CEN4 and the PCR signal was analyzed on 2% agarose gel.

To alleviate these problems we applied an alternative approach to concentrate centromeric nucleosomes using chromatin immunoprecipitation (ChIP) with anti-histone 3 (H3) N-terminal (H3-N) antibody, which detects N-terminal of H3 histone (Materials&Methods 2.3.3). Since Cse4 protein has an N-terminal domain with a unique sequence that does not share any similarities with H3 protein, centromeric nucleosomes containing Cse4 will not be detected by H3-N antibody and hence, could be enriched in the unbound samples. WB analysis demonstrated that the unbound sample does not contain H3 (**Figure 74**), indicating that we were able to deplete majority of H3 histones and thus, remove canonical nucleosomes from the soluble chromatin. We next sought to check the presence of centromeric nucleosomes with a standard PCR using 5´CEN4 and 3´CEN4 primers (**Figure 74**). While the bound sample lacks the centromeric DNA signal, the unbound sample actually shows a centromeric nucleosome band, showing that we managed to enrich centromeric nucleosomes and deplete canonical nucleosomes with our approach.

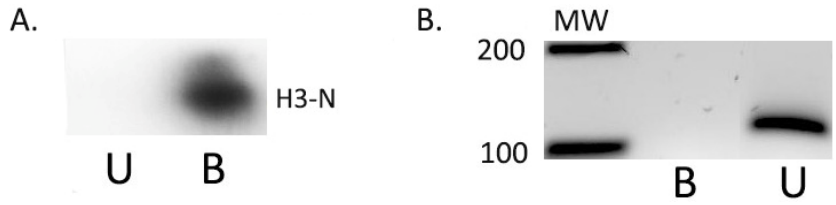


Figure 74: ChIP efficiency determined by western blot (WB) and PCR. A. H3 containing chromatin is detected by WB when crosslinked and MNase digested chromatin is subjected to immnuoprecipitation with antibody against H3-N. B. Presence of centromeric nucleosome is determined by standard PCR using 5´CEN4 and 3´CEN4. U stands for unbound material after ChIP and B for bound material.

Next, we layered the unbound sample containing centromeric nucleosomes on a 5-15% sucrose gradient. For simplicity, each two fractions were pooled and analyzed together. The presence of centromeric nucleosome was inspected by PCR. Centromeric DNA is mostly enriched in fractions 7, 8, 9 and 10 and slightly in fractions 5 and 6. In an attempt to elucidate the centromeric nucleosome organization, we then precipitated the protein contents of each fraction and analyzed them by 4-20% precast SDS-PAGE gel, followed by silver staining (**Figure 75**). Accordingly, the fragments containing centromeric nucleosomes (fractions 7-9) have an extra band around 25 kD, which might correspond to the 26-kD Cse4. Interestingly, these fragments display distinct histone protein contents. While in the top fractions the

histone bands H3, H2A/B and H4 are visible, centromere containing fractions lack H3 band, confirming Cse4 substitution for H3, although those fractions contain H2B, H2A bands and a faint H4 band **Figure 75**&Appendix 10). These findings demonstrate that the centromeric nucleosome composition is distinct from canonical nucleosomes and exclude the hexasome model that involves two molecules of Scm3 together with Cse4/H4 tetramer.

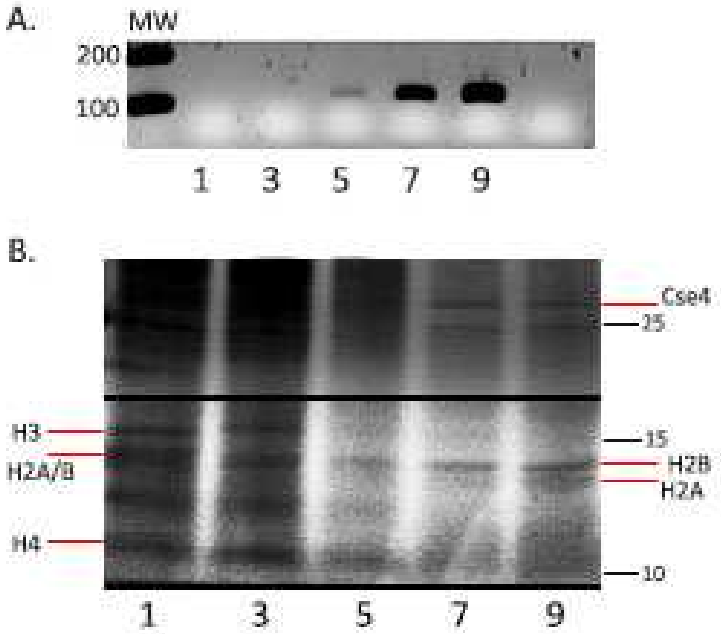


Figure 75: Characterization of centromeric nucleosome. A. The centromeric nucleosome enrichment in fractions from sucrose gradient is determined by PCR. B. The protein content of each fraction is visualized by Silver Staining of 4-15% SDS-PAGE gel.

However, for an accurate protein characterization, the bands should be further analyzed by quantitative mass spectrometry. Moreover, further experiments are required to elucidate the exact composition of centromeric nucleosomes, since our methodology is not able to differentiate octameric nucleosomes from hemisomes containing a single copy of each histone.

Finally, in an attempt to interrogate the centromeric nucleosome dynamics along cell cycle, especially at S phase, we inspected the presence centromeric nucleosome signal at different periods of cell cycle. For that reason, we isolated chromatin at late G1 phase and different points of S phase, digested with MNase and loaded on 5-15% sucrose gradient. The collected fractions were then examined by PCR to compare the centromeric

nucleosome signals (**Figure 76**). While we detect centromeric signal at S phase fraction 2 (25', 30' and 35' chromatin), at 0' lacks any centromeric signal and has a faint band at fraction 3. Moreover, the bands are more intense at S phase. This observation illustrates that centromeric nucleosome molecules are lighter at S phase, when Cse4 is recruited to centromeric DNA and replaces the old Cse4, which might also explain the more intense signals at S phase. The decrease in the density of centromeric nucleosomes at S phase could be due to absence of some proteins that stabilize centromeric nucleosomes, like Ndc10 or Smc3 (Cho & Harrison 2011). These proteins might bind to centromeric DNA after Cse4 recruitment, but would be absent at S phase. Based on these findings, the composition of centromeric nucleosomes seems to fluctuate along cell cycle.

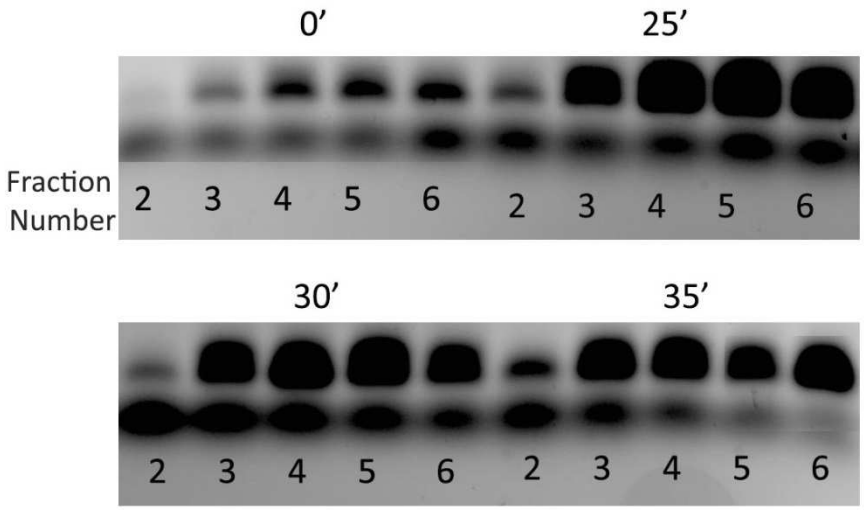


Figure 76: Centromeric nucleosome dynamics along cell cycle. The presence of centromeric nucleosomes was determined by standard PCR using 5´CEN4 and 3´CEN4 and the PCR signal was analyzed on 2% agarose gel.

DISCUSSION

The aim of this thesis is the understanding of nucleosome positioning mechanisms, nucleosome dynamic along cell cycle and the impact of nucleosome dynamics on gene expression regulation.

1 Genome-Wide Nucleosome Positioning and Its Determinants

We first analyzed genome-wide nucleosome positioning in *S. cerevisiae*. For that purpose, we generated genome-wide nucleosome maps based on two methodologies: deep sequencing (MNase-Seq) and tiling microarray (MNase-CHIP).

Our comparative analysis of the nucleosome maps generated from either MNase-Seq or MNase-CHIP technologies reveals that the majority of nucleosome-depleted and -enriched regions are conserved. However, even though the general pattern is very similar, there is some dissimilarity among the maps. Firstly, we have detected fewer nucleosome peaks in MNase-CHIP maps. Moreover, MNase-Seq map exhibits sharper peaks and wider 5' NFRs, whereas MNase-CHIP nucleosome peaks are less defined and 5' NFRs are narrower around TSSs and TTSs. Lastly, the regularly spaced array organization of TSS downstream nucleosomes is not observed in MNase-CHIP nucleosome maps.

One of the reasons that might account for the differences in number and shape of nucleosome peaks in different experiments is the distinct mapping resolution achieved in each technology. As MNase-Seq theoretically is able to map at single base-pair resolution, the detection of nucleosomes is more precise and in turn, results in better-defined peaks. Moreover, MNase-CHIP only maps the regions that are covered by array probes, due to probe dependency. Furthermore, different nucleosome identification strategies applied in each methodology might also contribute to nucleosome map deviations. While in MNase-Seq, nucleosome positioning is determined by sequencing of nucleosomal DNA, in MNase-CHIP, nucleosome depleted and enriched regions are determined by the ratio between nucleosomal and genomic DNA. Therefore, the genomic DNA profile might further bias the nucleosome profiles, since genomic signal subtraction might lead to over- or under-representation of nucleosome signals. Lastly, MNase-CHIP may over-

estimate the uniformity and density of nucleosomes, which could explain the strong positioning detected in downstream nucleosomes. Overall, these findings show that the MNase-Seq technology provides higher resolution, lacks probe dependency, enables more accurate analysis of relevant non-repetitive regions and a nearly whole-genome coverage, which has led us to base the majority of our study on MNase-Seq derived nucleosome maps.

We demonstrated that nucleosome maps show a non-uniform positioning along the yeast genome and exhibit a canonical nucleosome architecture around transcription start sites (TSSs), consisting of a 5' NFR flanked by two well-positioned nucleosomes (Yuan et al. 2005; Lee et al. 2007; Mavrich et al. 2008; Kaplan et al. 2009). TSSs are located within the first nucleosome peak 5' downstream, the so called +1 nucleosome. Notably downstream nucleosomes have higher signals than upstream nucleosomes around TSSs, indicating a higher nucleosome occupancy in genic than intergenic regions. On the other hand, the nucleosome profiles around transcription termination sites (TTSs) are more poorly defined, where downstream TTSs also known as 3´NFRs are highly depleted of nucleosomes, while upstream TTSs and downstream 3´NFRs are highly occupied. However, since genes are very close to each other over the yeast genome, downstream 3' NFR high signals might already correspond to the 5' signal of the adjacent gene in some cases.

We next sought to explore the role of DNA physical properties in nucleosome positioning around TSSs and TTSs as a potential determinant of non-random nucleosome positioning. Indeed, *cis*-acting factors, determined in terms of DNA sequence and its derived geometrical parameters, have long been considered to be one of the major players in nucleosome positioning. Crystal structures of nucleosome core particles revealed a lack of direct readout (i.e. nucleotide based) mechanism between histones and nucleosomal DNA (Luger et al. 1997; Richmond & Davey 2003). Furthermore, DNA affinities for nucleosome assembly preferences (such as the high-affinity Widom601 sequence) (Lowary & Widom 1998) are based on an indirect readout mechanism (Tolstorukov et al. 2007; Trifonov & Sussman 1980; Rohs et al. 2010; Xu & Olson 2010). Indirect evidences highlight the connection between DNA physical properties and chromatin organization, like unusual stiffness properties in human promoters (Goñi et al. 2007; Goñi et al. 2008; Miele et al. 2008). Therefore, a combination of direct and indirect readout

mechanisms appears to be required to achieve the correct interaction affinity and specificity.

Our analysis of MNase digestion profiles of genomic and nucleosomal DNAs revealed that MNase cut sites exhibit very similar nucleotide tetramer compositions, with a central d(A-T) step. On the other hand, such preferences are not observed in an equivalent control genomic DNA sample that was fragmented by sonication, indicating a strong MNase sequence preference that is intrinsic to genomic DNA, as reported earlier (Flick et al. 1986; Hörz & Altenburger 1981). Based on molecular dynamics (MD)-derived physical property analysis, these preferred MNase cut sites are characterized by a major groove wide opening and high flexibility that facilitates MNase access to those DNA regions, independently of the chromatin organization. Furthermore, MNase degraded regions, called low coverage regions (LRs), are very similar in genomic and nucleosomal DNA and are defined as highly flexible regions flanked by stiff motifs, as shown by dinucleotide-based stiffness profiles (**Figure 25**). Interestingly, in both genomic and nucleosomal coverage profiles, MNase-resistant sites are typically at the beginning of genes and upstream TTSs, while MNase-sensitive sites are mostly located upstream TSSs or downstream TTSs, defined as 5' and 3' NFRs (**Figure 26**). However, based on the coherence of nucleosome profiles, which are corrected by genomic DNA, we rule out a potential MNase bias. Indeed, the common nucleosome-depleted sites in both naked and nucleosomal DNA, 5' and 3' NFRs, are marked by particular helical parameters, indicating point flexibility flanked by stiff motifs and high deformation energy, prompting these regions to be more difficult to wrap around a histone octamer. These findings highlight the interplay between those intrinsic properties facilitating MNase digestion and those precluding nucleosome formation. In concordance with these observations, we detected anti-correlation between deformation energy and nucleosome coverage profiles around TSSs.

Taken together, our genome-wide MNase-digestion profile analysis demonstrates that 5' and 3' NFRs key regulatory regions are actually signaled by a differential pattern of MNase susceptibility in genomic DNA and intrinsic DNA physical properties determine major nucleosome signals in these regions. The unusual physical properties at TSSs and TTSs disfavor nucleosome positioning and hence, assist DNA access to regulatory proteins such as for instance, the accessibility of 5' NFRs to transcription machinery.

However, this should not involve constitutive gene expression, since in that case, there are other factors playing a role in gene regulation and nucleosome positioning, like chromatin remodelers and transcription factors. Therefore, DNA physical properties can only confer a basal mechanism of gene regulation.

To gain further insights into nucleosome positioning determinants, we compared the nucleosome profiles of genes that are regulated by distinct promoter types: TATA-box and TATA-less promoters. As reported earlier (Basehoar et al. 2004; Tirosh & Barkai 2008), nucleosome positioning differs between promoter types. TATA-containing genes, regulated by SAGA, tend to have higher transcription plasticity and fuzzier nucleosomes. On the other hand, TATA-less promoters, regulated by TFIID, usually correspond to housekeeping genes and have more well-defined nucleosomes. Moreover, TSS nucleosomes have lower occupancy and 5´NFRs are partially more occupied in TATA-containing promoters as compared to TATA-less promoters (**Figure 19**). The variations in nucleosome profiles between TATA-box and TATA-less genes can be mostly explained by the differential regulatory mechanisms. Genes with low transcription plasticity, i.e. TATA-less, are known to contain regulatory elements and TFBSs within 5' NFRs (Tirosh & Barkai 2008; Yuan et al. 2005). Therefore, TF binding at NFRs might not require chromatin remodeler action and subsequent disruption of nucleosomes. For that reason TATA-less genes exhibit well-phased nucleosomes that are barely disrupted upon gene activation. Moreover, since TFs usually bind to 5' NFRs, they are highly nucleosome depleted. Conversely, TATA-box genes possess more dynamic nucleosomes and their TFBSs are usually located within nucleosome occupied regions (Tirosh & Barkai 2008). Thus, upon gene activation, there is a competition between TFs and nucleosomes, which are disrupted by chromatin remodelers. As the transcription status of the TATA-box genes varies, nucleosomes are more dynamic, resulting in less-phased positioning. It is also worth mentioning that, TATA-box genes, which comprise 20% of yeast genes, present a noisier average nucleosome profile around TSSs due to the lower size sampling.

Previous studies assume that there is a well-established nucleosome distribution for a given species, which can be explained by intrinsic (DNA properties) or extrinsic (*trans*-acting) factors. However, our studies demonstrate a substantial dependence of the maps on experimental details,

which could suggest intrinsic fuzziness and variability in the nucleosome distributions that can be more stochastic than anticipated. Therefore, due to the variations in nucleosome positioning, the development of predictive models and exploring the determinants is very difficult and inadequate. For that reason, our next level of analysis involved the minimization of the variations within and among nucleosome maps. We inspected the extrinsic factors, derived from experimental conditions, data processing or heterogeneity of the samples, which may introduce variation among nucleosome maps.

We first explored the variation among nucleosome maps related to MNase digestion levels. To investigate its effect on nucleosome positioning, we compared nucleosome profiles of over- and under-digested samples against regularly digested samples. Over-digested derived nucleosome maps exhibit an increase in fuzzy and missing nucleosomes, especially -1 nucleosomes, besides a higher mean deviation of the nucleosome dyad position. These observations indicate that excessive MNase digestion leads to intra-nucleosome cleavage or complete disassociation of unstable nucleosomes. On the other hand, under-digested maps yield longer fragments, such as dinucleosomal DNA, which can introduce another type of noise, since their dyad alignment can result in counter-phased positioning with respect to mononucleosome signals, leading to higher noise especially in the linker region. Therefore, MNase digestion levels appear to strongly influence nucleosome maps and should be taken into account in comparative analysis.

We then investigated the putative variations arising from cell heterogeneity, since most of the nucleosome maps are derived from asynchronous cultures. We compared the nucleosome profiles around TSSs generated from unsynchronized and synchronized cultures. In asynchronous derived maps, there is a general increase in -1 and +1 nucleosome fuzziness, especially for cell-cycle dependent genes, indicating that asynchrony might produce considerable amount of noise due to cell cycle-dependent gene expression variations, chromatin compaction or relaxation along cell cycle. Then, the nucleosome maps generated from asynchronous cultures actually reproduce average nucleosome phasing of a cell population at different cycle stages, which may particularly bias those regions with striking cell-cycle dependent chromatin alterations. Interestingly, despite the changes in nucleosome

phasing under distinct conditions, 5´NFRs are conserved regardless of experimental conditions or cell heterogeneity.

Based on our findings, we subsequently attempted to minimize the noise due to asynchrony and MNase digestion levels in order to explore the intrinsic nucleosome positioning variations. For that purpose, we compared the nucleosome maps derived from two late G1 phase-synchronized replicas. Interestingly, the nucleosome organizations are quite similar around TSSs. The majority of genes displays well-positioned -1 and +1 nucleosomes with either open (wider) or closed (narrower) NFRs. However, individual gene analysis exhibits clear differences, which are mostly subtle and affecting the position of only one nucleosome, such as changing from W (well-positioned) to F (fuzzy). Indeed, only 3% of the nucleosomes are located exactly at the same position along the genome in both replicas, pointing out the intrinsic nucleosome dynamics. Therefore, the well-positioned nucleosomes might not be tightly wrapped, but are mobile along genome, since intrinsic sliding barriers are very small.

Finally, in order to study the nucleosome positioning determinants more accurately, we minimized the noise due to asynchrony, inter-replica variations and MNase digestion levels by selecting a set of genes from late G1 phase arrested samples that show high correlation in nucleosome profiles between replicas and performed our further analysis with those genes. We demonstrated that once NFR is defined, as open or closed, a simple statistical model can predict the nucleosome positioning around TSSs, where nucleosome arrays are located starting from NFRs with a lineal decay in positioning. Deformation energy computed around gene TSSs, particularly those with open NFRs, demonstrates that NFRs are marked by higher deformation energy, indicating that physical properties can define the boundaries of NFRs, confirming our previous results. On the other hand, the central NFRs display lower deformation energy values, but are enriched in predictive TFBS signals. Thus, even though nucleosome formation is allowed in the middle of NFRs in terms of energy cost, nucleosome binding is avoided due to the competition with TFs. In summary, DNA physical properties and TFs act synergistically to define NFRs and hence the nucleosome arrays, which might also explain the robustness of NFRs to experimental variations. This synergetic effect of DNA physical properties and TFs on nucleosome positioning is also observed in promoters with closed NFRs, displaying higher

deformation energy and predictive TFBS signal at NFRs. However, the boundaries of NFRs are less defined in terms of deformation energy, which is also higher where -1 nucleosome is positioned. Perhaps, in the closed promoters, nucleosome positioning is more chromatin remodeler dependent, like TATA-box promoters, which might replace -1 nucleosome despite the high energy barrier. On the other hand, away from 5' NFRs, nucleosomes are highly dynamic, changing their location constantly due to the absence of a high energy barrier. Probably, chromatin remodelers play the major role in positioning of such nucleosomes. Notably, the effect of cis- and trans-acting factors on nucleosome positioning is lost in *in vitro* reconstitution experiments, which may explain the differences between *in vivo* and *in vitro* nucleosome maps.

Overall, our study provides a detailed genome-wide analysis of nucleosome maps in *S. cerevisiae* in order to explore the possible nucleosome determinants. We were able to demonstrate that in yeast, promoter regions are in general marked by unusual physical properties. This, in turn, results in high energy barriers at 5´NFRs, where most of TFs are bound. Therefore, nucleosome formation at 5´NFRs is highly prevented by a synergetic effect of DNA physical properties and TFs. Nucleosomes are then positioned against the 5´NFR barrier by the help of other factors, like chromatin remodelers with a lineal decay in phasing.

2 Chromatin Dynamics throughout Cell Cycle

During eukaryotic cell cycle, chromatin undergoes several conformational changes, i.e. decondensation in S phase or compaction during chromosome segregation. In this regard, we aimed at analyzing cell-cycle dependent chromatin dynamics based on nucleosome maps derived from synchronized cultures at each cell cycle phase.

The comparison of nucleosome maps and MNase digestion patterns at each cell cycle phase reveals that chromatin at S phase displays fuzzier nucleosomes and higher sensitivity to MNase digestion along yeast genome as compared to G1, G2 and M phases (**Figure 46).** This differential sensitivity indicates higher chromatin disorganization at S phase that might account for DNA replication. Conversely, chromatin at M phase appears to be the most

resistant to MNase digestion, indicating higher compaction, in agreement with mitotic condensation of chromatin. G1 and G2 phases display a very similar compaction level, which is lower than M phase. Furthermore, since DNA replication is originated from multiple sites at different time points, we studied chromatin organization from a time-course experiment at short frame intervals within the S phase, which turned to be the most prone to MNase digestion, confirming previous observations. Yet at a particular point during S phase, chromatin shows the highest MNase sensitivity, which decreases gradually; indicating a step-wise chromatin condensation/decondensation along cell cycle. This observation was further confirmed by Bioanalyzer derived profiles of typically digested chromatin extracts, where a particular sample at S phase shows the highest fragmentation level and enrichment in half-nucleosome sized fragments (**Figure 49**).

Based on our nucleosome map analyzes, nucleosome profiles around TSSs reveal an anti-correlation between MNase sensitivity and nucleosome occupancy or phasing. The samples displaying higher MNase sensitivity, i.e. those collected at S phase, show lower occupancy and phasing around TSS nucleosomes. Particularly, our gene classification analysis demonstrates that S and M phases display an increase in fuzzy -1 and +1 nucleosomes. Furthermore, nucleosome coverage standard deviation (SD) and genome-wide coverage comparison confirm the global fuzziness at S phase, although M phase exhibits fuzziness along the genome to a less extent. Of note, one of the NGS technology limitations is the inability of small fragment detection, which might mask their effect on the fuzziness and occupancy of the nucleosome maps and in turn, precludes a gradual chromatin change evaluation.

Taken together, we demonstrated that chromatin at S phase exhibits a distinct conformation, showing higher sensitivity to MNase and displaying fuzzier nucleosomes along yeast genome. These findings are especially striking at a particular S phase point, indicating that the conformational changes are sudden and acute. In light of this, perhaps this particular period corresponds to the activation peak of most of the replication origins and hence, the majority of chromatin should be then disassembled to allow DNA replication initiation. Another possible scenario implies that replicated DNA still maintains high acetylation levels that might lead to more loosen

nucleosomes, since histone acetylation weakens histone DNA interactions at the entry and exit points of nucleosomes, being more dynamic and fuzzier, as demonstrated earlier (Masumoto et al. 2005).

On the other hand, chromatin at G1 phase displays the highest phasing around TSSs and along the genome, which is followed by G2 phase. Since G1 and G2 phases are periods of cell growth and protein synthesis, chromatin organization is highly regulated. Therefore, at these stages nucleosomes have higher phasing and occupancy.

Lastly, we would expect that chromatin compaction at the M phase should increase nucleosome positioning, as the degrees of freedom become more limited due to interchromatin contacts. However, we observed higher degree of fuzziness around TSSs and along yeast genome, as compared to G1 and G2 phases. One possible explanation for the phasing decrease might be the absence of a 30 nm fiber organization. As indicated previously in human cell lines (Nishino et al. 2012), yeast mitotic chromatin might predominantly involve irregularly folded nucleosome fibers, with an increase in flexibility and dynamics of nucleosomes, implying higher DNA access. Moreover, the fact that the cells collected at M phase already start losing the synchrony after alpha factor release might contribute even more to nucleosome fuzziness, like in the case of asynchronous culture.

We next sought to explore the cell-cycle dependent nucleosome fluctuations at individual genes. The analysis of nucleosome profile transitions between cell cycle stages demonstrates that nearly half of the genes maintain the well-positioning of -1 and +1 nucleosomes, indicating that only certain genes are affected by cell cycle chromatin dynamics. Some genes changing their nucleosome pattern around TSSs usually have well-positioned nucleosomes at G1 and G2 phases, which become fuzzy at S and M phases. Another set of genes show less phasing at G1 phase and are usually involved in response to pheromone, conjugation and sexual reproduction. This observation indicates that despite the overall high nucleosome phasing at G1, some genes display fuzzy nucleosomes due to an increase in expression. On the other hand, NFRs are quite conserved and maintain their “open” or “closed” status, suggesting that NFRs are not affected by cellular processes along cell cycle. Genes with a closed NFR state are highly enriched in TATA-box, while those with open NFR state are mainly TATA-less genes, in agreement with previous studies (Tirosh & Barkai 2008; Zaugg & Luscombe 2011). and as discussed earlier.

Interestingly, open NFR state genes tend to be under control of chromatin remodelers INO80 and ISW1, which are shown to act collectively over many positions on gene bodies (Yen et al. 2012). Nucleosome positioning analysis from deletion strains also demonstrated that INO80 and ISW1 remodelers promote shifts in downstream and upstream nucleosomes away from NFRs (Tirosh et al. 2010; Yen et al. 2012), resulting in wider NFRs. Therefore, in open state NFR genes a high energy barrier at NFRs may cause a directional shift in the position of nucleosomes away from NFRs under control of INO80 and ISW1 and these remodelers can position nucleosomes against a barrier in a regularly spaced array.

In addition, we aimed to decipher the interplay between chromatin organization and transcription by using gene expression data at different cell cycle phases. Even though we were not able to find a general trend between nucleosome positioning and gene transcription level, we observed clear differences in average nucleosome profiles between highly and lowly expressed genes. At higher transcription rates, nucleosome occupancy decreases around promoter regions, which is likely due to RNA polymerase II (RNAPII) binding to promoters, leading to nucleosome disassembly. Notably, inactivation of RNAPII results in an increase in -1 nucleosome occupancy (Weiner et al. 2010). However, coding region nucleosomes display higher occupancy at increased transcription levels, which might be explained by the requirement of an ordered nucleosome structure as an additional control mechanism to facilitate RNAPII movement along DNA. Therefore, high nucleosome occupancy at the less ordered S and M phases assures an even more tightly controlled transcription of highly expressed genes. Moreover, the same trend observed in nucleosome coverage and expression differences between subsequent phases, i.e. the lowest changes of both coverage and expression occur at S-G2 phase transition, indicates that some nucleosome coverage changes might be due to changes in gene expression along cell cycle and further supports the interplay between transcription and chromatin.

Deeper analysis of cell-cycle dependent genes illustrates the complexity of the relationship between gene expression and nucleosome positioning. The transcription dependent changes in nucleosome positioning do not occur uniformly in the genes. Upon activation, most of the genes, especially alpha-factor pathway genes, show a dramatic decrease in nucleosome occupancy

around TSS. Other genes display a shift in positioning of TSS flanking nucleosomes, while some genes clearly have a distinctive pattern when their expression is lower. Usually, genes that are similarly regulated also show similar nucleosome profile variations. Conversely, some cell-cycle regulated genes maintain their nucleosome profile along cell cycle, despite the fluctuation in their expression, implying a non-uniform mechanism for transcription-dependent chromatin regulation or *vice versa*.

Overall, these findings suggest that the interplay between chromatin organization and transcription is more complex than as anticipated before. There are various factors playing a role in transcription, like histone exchange, histone modifications and noncoding RNAs might also affect chromatin organization (Rando & Winston 2012), perhaps at a level different than simple nucleosome organization.

Finally, we studied nucleosome positioning around DNA replication origins along cell cycle. Previously, genome-wide nucleosome positioning revealed that origin recognition complex (ORC) binding sites, the so-called ARS consensus sequence or ACSes, are nucleosome depleted while surrounding regions are occupied by well-positioned nucleosomes (Eaton et al. 2010; Berbenetz et al. 2010). In accordance with this, our comparative nucleosome positioning analysis around replication origins reveals a similar pattern to that of TSSs, although origin-NFRs are flanked by well-positioned nucleosomes in a symmetrical manner. We further observed that nucleosome profiles around replication origins are quite conserved along cell cycle, due to the constitutive binding of ORC to ACSes along cell cycle. Then, ORCs might act as a barrier around replication origins against which nucleosomes are packaged symmetrically. Interestingly, replication origins at G1 phase exhibit overall lower occupancy and higher phasing in NFR flanking nucleosomes and slightly wider NFRs around origins. This might be explained by the recruitment of other replicatory proteins, i.e. MCM complex, to form the pre-Replication complex, which probably leads to preclusion of nucleosome accessibility to longer DNA segments, resulting in wider NFRs. The protein binding might additionally lead to disassociation of flanking nucleosomes to increase DNA accessibility to the replication complex, resulting in a decrease of nucleosome occupancy at G1 phase. However, once origins are activated at S phase and the complex disassociates, nucleosomes might slide towards origin-NFRs. Therefore, ACSes after G1

phase can be partially occupied. However, due to constitutive binding of ORCs, as mentioned, the differences in NFR width along cell cycle are marginal.

Furthermore, the diversity in replication origin structure is highlighted by individual origin profile analyzes, as reported before (Berbenetz et al. 2010). Replication origins show a remarkable broad NFR width ranges, from 80 to 600 bp. Based on these findings, our next goal was to analyze whether variations in chromatin structure are responsible for variations in origin activity. We were unable to find any relationship between replication origin NFR width and timing, but DNA physical properties. This observation led us to speculate that the intrinsic DNA properties might rather influence nucleosome organization around origin. However, phasing and occupancy of ACS nucleosomes were higher at early firing origins, which in turn, contain slightly higher H2A.Z levels. Thus, early origin activation might be more nucleosome-dependent and have more ordered chromatin organization, which is partially encoded epigenetically by H2A.Z-containing nucleosomes. On the other hand, the activity of lately firing replication origins, some of which are non-functional, could be more nucleosome-independent, as they have a more disorganized chromatin structure. Yet, to further refine our knowledge about the connection between origin activity and chromatin structure, we should include replication origin timing data in our nucleosome positioning analysis. Furthermore, the classification of nucleosome profiles and origin timing is necessary to improve a detailed analysis.

3 Centromeric Nucleosome Organization

Several studies have reported that centromeric chromatin has different features compared to bulk chromatin (Dunleavy et al. 2005; Verdaasdonk & Bloom 2011; Henikoff & Dalal 2005). However, the exact composition and structure of centromeric nucleosome in *S. cerevisiae* is still under debate. To this end, we tried to explore the centromeric nucleosome organization under physiological conditions.

Our nucleosome positioning analysis at centromeres discloses a perfect positioning of centromeric nucleosomes, meaning that all of the centromeric nucleosomes within the populations are located exactly at the same genomic

place, i.e. CDE regions, and are not dynamic unlike canonical nucleosomes. These findings confirm the strong sequence-dependent recruitment and positioning of centromeric nucleosomes.

The absence of some centromeric nucleosome signals in our nucleosome maps raised the concern about the undertaken experimental protocols, which might not be optimum for centromeric nucleosome isolations. To overcome this potential issue, we used a modified protocol andincluded sucrose gradient fractionation after chromatin immunoprecipitation (ChIP). Very interestingly, centromeric nucleosomes tend to be heavier than canonical mononucleosomes and might be even as heavy as dinucleosome particles. Therefore, centromeric nucleosomes are not likely to adopt a hemisome structure, only containing a single copy of each histone, which was previously suggested (Dalal et al. 2007). However, this assumption relies on the absence of any additional proteins, which otherwise, might also contribute to the apparently heavier structure of centromeric nucleosomes. Furthermore, we also determined the protein content of centromeric nucleosomes, which possess clear bands of H2A and H2B proteins and a faint H4 band. Intriguingly, H3 protein band is absent, indicating that in the centromeric nucleosome organization, Cse4, H2A, H2B and H4 proteins are involved; excluding the hexasome model that contains Cse4/H4 tetramer together with Scm3 proteins. Altogether, our experiments favor octameric nucleosome organization in centromeres under physiological conditions. However, future experiments are on going to confirm the exact composition of centromeric nucleosomes, such as mass spectrometry analysis.

Finally, we also explored the centromeric nucleosome dynamics along cell cycle based on our genome-wide nucleosome maps. However, because of potential bias due to experimental procedures, our analyses are not conclusive. Nevertheless, sucrose gradient density experiments demonstrate that centromeric nucleosomes have a lighter configuration at S phase, when their recruitment and deposition takes place. This might be explained by the absence of certain proteins, which are not bound to centromeric nucleosomes for their stabilization at S phase yet. Another possible explanation could be the centromeric nucleosomes display an intermediate configuration during DNA replication and Cse4 deposition, which later transit to octameric organization.

Overall, we show that Cse4 containing nucleosomes have a distinct nucleosome organization and positioning. Furthermore, we proposed that their composition might oscillate along cell cycle and intermediate forms might exist. However, to confirm our hypothesis further experiments are needed to determine protein contents along cell cycle by using alternative approaches.

CONCLUSIONS

5’ and 3’ NFRs key regulatory regions show different MNase sensitivity and are signaled by differential intrinsic DNA physical properties.

1. Nucleosome formation at 5´NFRs is highly prevented by a synergetic effect of DNA physical properties and *trans*-acting factors. Once NFR is defined, a simple statistical model can predict the nucleosome positioning around TSSs, starting from NFRs with a lineal decay in positioning.

2. Chromatin organization varies along cell cycle. G1 and G2 phases have a more organized chromatin, while S and M phases are characterized by fuzzier nucleosomes and lower nucleosome occupancy in promoter regions. A high chromatin disorganization is especially striking at a particular S phase point.

3. Nucleosome distribution around TSSs and most likely along the entire genome is quite plastic, where nucleosomes move in a one-dimensional space until a strong depletion signal (NFR) that helps nucleosome fiber organization by means of simple statistical positioning.

4. Even though there is not a global relationship between nucleosome organization and gene expression, highly expressed genes display different nucleosome organization compared to lowly expressed ones. Moreover, genes displaying the largest nucleosome profile changes are also showing high expression plasticity and are usually cell-cycle-regulated genes.

5. Nucleosome profiles around replication origins are quite conserved along cell cycle. However, G1 phase chromatin exhibits overall lower occupancy and higher phasing in NFR flanking nucleosomes and slightly wider NFRs around origins. Moreover, nucleosome positioning differs between replication origin types: early firing origins display higher phasing and occupancy in nucleosomes around ACSes, as compared to late firin origins.

6. Cse4 containing nucleosomes have a distinct nucleosome composition and positioning. Our study supports octameric nucleosome organization in centromeres under physiological conditions. Furthermore, nucleosome composition might oscillate along cell cycle with intermediate forms.

SUMMARY IN SPANISH

Introducción:

El nucleosoma es la unidad estructural de compactación del ADN en organismos eucariotas y está formado por el enrollamiento de 147 pares de bases (pb) de ADN de cadena doble alrededor de un octámero de proteínas llamadas histonas. El posicionamiento de nucleosomas regula el genoma eucariota, ya que controla la accesibilidad del ADN a las proteínas reguladoras, y por tanto, influye en muchas funciones celulares como la transcripción, la replicación y la reparación del ADN (Luger et al. 1997; Felsenfeld 1992).). Así pues, la determinación de la posición exacta de los nucleosomas es clave para comprender los fundamentos de procesos celulares complejos. Hasta la fecha, gracias al rápido desarrollo en técnicas de secuenciación de alto rendimiento, se han elaborado mapas de nucleosomas a alta resolución en levadura, gusanos, moscas y humanos (Albert et al. 2007; Mavrich et al. 2008; Schones et al. 2008).). En general, los mapas genómicos de nucleosomas derivados de distintos organismos muestran que la mayoría de nucleosomas están "bien posicionados" o "en fase", lo cual significa que se sitúan exactamente en el mismo lugar del genoma en todas las células que componen una población. Concretamente, en *S. cerevisiae*, el 80% de los nucleosomas están bien posicionados y separados entre sí por 18 pb de DNA enlazador (Yuan et al. 2005; Mavrich et al. 2008; Lee et al. 2007). Por otro lado, algunos nucleosomas, especialmente en las regiones intergénicas, se encuentran peor posicionados y nos referimos a ellos como nucleosomas “borrosos” (“fuzzy”, en inglés). Éstos ocupan distintos lugares en el genoma de una población celular, lo que sugiere que los nucleosomas pueden colocarse de forma dinámica a lo largo de las hebras de ADN (Tanaka et al. 1996; Fragoso et al. 1995). Asimismo, estudios a nivel de genoma revelaron que el posicionamiento de nucleosomas no es al azar, sino que sigue un patrón canónico alrededor del lugar de inicio de transcripción (TSS, por sus siglas en inglés), el cual consta de una región libre de nucleosomas (NFR, por sus siglas en inglés) flanqueada por nucleosomas bien posicionados. Se ha sugerido que la conservación de este patrón general es crucial para la regulación de la expresión génica.

Existen diferentes mecanismos que determinan el posicionamiento de los nucleosomas. Intensos esfuerzos de investigación revelaron que el posicionamiento de nucleosomas no está determinado solamente por un factor, sino más bien por la combinación de muchos factores, tales como las

propiedades intrínsecas del ADN, los remodeladores de la cromatina dependientes de ATP, las modificaciones postraduccionales de las histonas, variantes de histonas o factores de transcripción.

La organización de la cromatina se mantiene o se altera según el estado celular mediante la inclusión, extracción o intercambio de nucleosomas. El ciclo celular es uno de los factores que determinan la estructura global de la cromatina y, por lo tanto, el posicionamiento de nucleosomas. La compactación de la cromatina se puede condensar o relajar dependiendo de la etapa específica del ciclo celular. De hecho, la progresión de este ciclo requiere la desintegración parcial de la estructura de la cromatina para regular el metabolismo del ADN, la cual se consigue por la interrupción de los contactos del ADN con las histonas a través de dos niveles principales: los factores de remodelación de la cromatina dependientes de ATP o las modificaciones postraduccionales de las colas de histonas, mencionadas anteriormente (Boeger et al. 2003). Especialmente durante la replicación, algunas interacciones específicas de ADN e histonas tienen que ser desensambladas y reensambladas para una replicación fiable del ADN y una estructuración adecuada de la cromatina.

En este estudio hemos analizado el posicionamiento de nucleosomas en *S. cerevisiae* a nivel genómico, utilizando técnicas de secuenciación masiva, para explorar los posibles factores determinantes de dicho posicionamiento. Por esta razón, hemos analizado la función de las propiedades físicas del ADN en la determinación del posicionamiento de nucleosomas mediante el análisis comparativo entre ADN genómico y nucleosomal. Mediante la selección de un conjunto de genes donde se minimizó el ruido experimental en los mapas de nucleosomas, hemos examinado la sinergia entre la unión de factores de transcripción y las propiedades físicas del ADN en la determinación de las NFR. Adicionalmente, para medir la movilidad intrínseca de los nucleosomas en el genoma a lo largo del ciclo celular, hemos analizado mapas de nucleosomas a distintos instantes del crecimiento celular en una población sincronizada de *S.cerevisiae.* Nuestros estudios muestran la interacción entre la expresión génica y la dinámica de los nucleosomas durante el ciclo celular. Más allá de los cambios globales, también se han estudiado con mayor detalle los orígenes de replicación y los genes dependientes del ciclo celular.

La estructura y composición del nucleosoma centromérico son motivo de controversia. Distintos estudios han propuesto modelos contradictorios acerca de la estructura precisa del nucleosoma centromérico. Entre los muchos modelos propuestos, existen tres más populares: hemisoma, octasoma y hexasoma. Finalmente, hemos tratado de determinar el contenido proteico del nucleosoma centromérico, así como su fluctuación a lo largo del ciclo celular.

Objetivos

Los objetivos de esta tesis son entender cómo se posicionan los nucleosomas, explorar la aportación de los factores *cis* como determinantes de este posicionamiento, comprender mejor la relación entre regulación de la expresión genética y la dinámica de los nucleosomas a lo largo del ciclo celular, alrededor de los TSS y de los orígenes de replicación, así como dilucidar la composición de los nucleosomas centroméricos.

Resultados

1. **Impacto de las propiedades físicas en el posicionamiento de los nucleosomas**

La secuencia del ADN se ha considerado un factor importante en el ensamblaje de nucleosomas Las estructuras cristalinas de nucleosomas revelaron una falta de mecanismos directos de lectura entre las histonas y las bases de ADN (Luger et al. 1997; Richmond & Davey 2003), lo que sugiere un posible mecanismo de lectura indirecto. Para entender el papel de las propiedades físicas del ADN en el posicionamiento de nucleosomas, hemos utilizado los perfiles de digestión con nucleasa micrococcal (MNasa) de ADN genómico y nucleosomal en levadura. En primer lugar, hemos comparado los sitios de corte de la MNasa, definidos como los tetrámeros formados por las dos bases anteriores y posteriores del lugar de corte, con tal de encontrar preferencias en la escisión del ADN. La MNasa muestra fuertes preferencias de secuencia en ambas muestras, aunque son más notables en el ADN genómico, observando patrones similares de corte entre el ADN genómico y el nucleosomal. A continuación, hemos comparado las regiones digeridas por MNasa causadas por su actividad exonucleasa. Se definieron esas regiones como lugares de baja cobertura (LR, del inglés *Low-covered Regions*). Observamos una gran similitud entre LR genómicos y

nucleosomales en cuanto a la composición de tetrámeros y así como en las regiones del genoma afectadas. Definimos a continuación la intersección de las LR en ambas muestras, dando lugar a las regiones comunes con baja cobertura (CLR, del inglés Common Low-covered Regions). Analizado las propiedades físicas derivadas de dinámicas moleculares, se observó que los sitios de corte preferente de la MNasa, así como las CLRs, se caracterizan por unas propiedades físicas inusuales. Por otra parte, regiones del ADN nucleosomal más resistentes a la actividad enzimática se concentran principalmente en el inicio y final de la región codificante de los genes. Las mismas regiones aparecen como regiones menos sensibles a la MNasa en el ADN genómico. Por otro lado, las regiones más sensibles a la MNasa tanto en ADN nucleosomal como genómico se detectan principalmente antes de los inicios de transcripción o después del término de transcripción, coincidiendo con las NFR 5' y 3' que se pueden observar en los mapas de nucleosomas, indicando que dichos mapas no sólo podrían reflejar el posicionamiento de nucleosomas, sino también la susceptibilidad intrínseca del ADN genómico a la digestión por MNasa. Para esclarecer el posible impacto de las propiedades intrínsecas del ADN en estas regiones reguladoras, hemos calculado la energía de deformación que se requiere para envolver una secuencia de ADN alrededor de un octámero de histonas, mediante una simple función de energía elástica basada en el descriptor físico derivado de dinámicas moleculares. Se encontró que, en general, las CLRs se correlacionan con una alta energía de deformación, confirmando que estas regiones son más rígidas para enrollarse alrededor de un núcleo de histonas. Por lo tanto, debido al alto coste energético, la formación de nucleosomas en esas regiones es menos factible. En general, la correlación entre las CLRs y la alta energía de deformación sugiere que propiedades intrínsecas, tales como la flexibilidad, induzcan a que un segmento de ADN que es buen sustrato para la actividad enzimática de la MNasa también evite su enrollamiento en torno a un nucleosoma.

2. Difusión y ruido en el posicionamiento de nucleosomas

Hasta la fecha se han generado mapas de posicionamiento de nucleosomas a nivel genómico en muchos organismos (Kaplan et al. 2008; Mavrich et al. 2008; Schones et al. 2008; Lee et al. 2007). Cada uno de estos mapas muestra una estructura de cromatina muy similar alrededor del TSS, aunque el posicionamiento individual puede diferir notablemente. De este modo,

nucleosomas que aparecen muy bien posicionados en un estudio, aparecen difusos o simplemente ausentes en otros (Huebert et al. 2012; Bai & Morozov 2010; Kuan et al. 2009; Tsankov et al. 2010).). La discrepancia entre los mapas de nucleosomas puede provenir de diferentes fuentes: i) las condiciones experimentales (como los niveles de digestión de MNasa o el protocolo de secuenciación); ii) el procesamiento de datos para la detección de nucleosomas, iii) la heterogeneidad de las muestras provenientes de diversos estados celulares en el cultivo; y iv) la dinámica de los nucleosomas en el genoma, que puede traducirse en picos de ocupación difusos (Belch et al. 2010; Lehner 2010). En un intento de medir el efecto del ruido en la ocupación y posicionamiento de los nucleosomas, se inspeccionaron los factores extrínsecos que pueden inducir dicha diversidad comparando mapas de nucleosomas obtenidos mediante MNasa-Seq en diferentes condiciones.

Con tal de minimizar el ruido producido por los protocolos experimentales, se compararon réplicas de mapas de nucleosomas deteniendo el ciclo celular en la fase G1 y utilizando secuenciación *paired-end*. En general, los patrones de nucleosomas aparecen similares en ambas réplicas. No obstante, a pesar de que el 90% de los genes muestran perfiles de cobertura similares, se observan algunos cambios sutiles, como la degradación en el posicionamiento de los nucleosomas -1 o +1, pasando de bien posicionado a difuso. Curiosamente, las NFRs no parecen cambiar entre réplicas, lo que indica que están más conservadas y que los nucleosomas que las flanquean son menos propensos a variaciones. Estos hallazgos muestran que las variaciones inter-réplicas son relevantes y señalan que el posicionamiento de nucleosomas es intrínsecamente plástico y dinámico.

Con el fin de determinar la variabilidad causada por la heterogeneidad celular, incluimos un cultivo de levadura no sincronizada en nuestro análisis y comparamos su perfil de nucleosomas con las muestras sincronizadas. Los cambios más importantes se encuentran en la eliminación gradual de los nucleosomas -1 y 1, que son más difusos en la muestra no sincronizada (asíncrona), sugiriendo que la reorganización de la cromatina en función del ciclo celular puede sesgar los experimentos no sincronizados de MNase-Seq. Por otra parte, el ancho de la NFR entre muestras sincronizadas y asíncronas permanece estable y muestra cambios similares entre réplicas biológicas. En general, los experimentos asíncronos contienen una fuente adicional de ruido debido a la dinámica de nucleosomas durante el ciclo

celular. Por lo tanto, es necesario tener especial cuidado con los mapas derivados de muestras asíncronas (la mayoría de los disponibles en la bibliografía) ya que estos mapas promedio pueden ocultar diferentes poblaciones con arquitecturas de nucleosomas completamente distintas. Además, para investigar el posible sesgo causado por una digestión diferencial de la MNasa, hemos utilizado dos experimentos MNasa-Seq adicionales derivados de un cultivo sincronizado en G1 pero tratados bajo condiciones más agresivas (sobre-digiriendo la muestra) o más suaves (digiriendo menos la muestra). La muestra sobre-digerida muestra un aumento de los nucleosomas -1 difusos o ausentes, sugiriendo que el exceso de digestión por MNasa puede conducir a una degradación parcial de algunos nucleosomas bien posicionados. Esto puede resultar en picos de nucleosomas difusos o incluso completamente degradados en el caso de los nucleosomas más inestables, comportando la pérdida de la señal del nucleosoma en esa posición. Por otro lado, la muestra con baja digestión presenta un incremento notable de las regiones con baja cobertura. Adicionalmente, en esta muestra también se observan en general fragmentos más largos, que podrían ser otra fuente de ruido en los mapas de nucleosomas ya que los algoritmos de posicionamiento de nucleosomas suelen considerar los picos observados como mono-nucleosomas y alinearlos sobre el genoma en función de su posición central, suponiendo que se corresponde con la díada del nucleosoma.

Para minimizar el ruido debido a la asincronía entre réplicas y los niveles de digestión de MNasa, hemos seleccionado un conjunto de genes a partir las muestras en G1 que muestren una alta correlación entre los perfiles de nucleosomas de ambas réplicas, y poder evaluar así nuestros modelos predictivos. La mayoría de estos genes presenta un buen posicionamiento en los nucleosomas -1 y +1 y una NFR abierta o cerrada. Hemos demostrado que una vez la NFR se define como abierta o cerrada, un modelo estadístico simple puede predecir el posicionamiento de nucleosomas en torno a los TSSs, donde se encuentran cadenas de nucleosomas con un decaimiento lineal en su posicionamiento. La energía de deformación calculada alrededor de los TSSs, en particular de los que tienen NFR abiertas, demuestra que las NFRs están señaladas por una mayor energía de deformación, indicando que las propiedades físicas pueden, en efecto, definir los límites de las NFRs. Por otro lado, las NFRs de apertura intermedia muestran valores de energía de deformación más bajos pero se encuentran enriquecidas en la predicción de

señales de sitios de unión de factores de transcripción (TFBS, por sus siglas en inglés). Este efecto sinérgico entre las propiedades físicas del ADN, las señales de los TFBS y el posicionamiento de los nucleosomas se observa también en los promotores con NFR cerrados, mostrando mayores energías de deformación así como señales de TFBS en la NFR. Sin embargo, los límites de estas NFR están menos definidos en términos de energía de deformación donde se coloca el nucleosoma -1.

3. Dinámica de la Cromatina a lo largo del Ciclo Celular

En las células eucariotas, la organización de la cromatina varía conforme el ciclo celular avanza, especialmente en la fase S, cuando tiene lugar la replicación del ADN, y en la fase M, cuando los cromosomas se condensan. Además, la expresión de aproximadamente 800 genes relacionados con el ciclo celular también se encuentra parcialmente modulada por la estructura de la cromatina en los promotores de estos genes. Con el fin de estudiar la relación entre la dinámica de la cromatina y la expresión de genes, así como variaciones dependientes del ciclo celular, hemos analizado el posicionamiento de nucleosomas a lo largo del ciclo celular usando poblaciones sincronizadas de *S. cerevisiae.*

En primer lugar, para ver el efecto general de la dinámica de la cromatina a lo largo del ciclo celular, diseñamos un experimento de digestión con MNasa recogiendo muestras en diferentes fases del ciclo. Hemos observado que la digestión por MNasa en la fase S muestra un patrón distinto en comparación con el de otras fases. En la fase S, los fragmentos digeridos son más uniformes y más cortos, lo que indica una mayor sensibilidad a la enzima. Con el fin de comprobar si la cromatina es más sensible a la MNasa a lo largo de la fase S, se incluyeron muestras adicionales recogidas a lo largo de la fase S. A pesar de que todas las muestras recogidas en la fase S muestran un mayor nivel de sensibilidad a la MNasa, un punto en particular, alrededor de los 30', muestra la mayor sensibilidad, causando una menor fragmentación de la cromatina. Para analizar la dinámica de la cromatina a lo largo del ciclo celular y cómo la organización en la fase S afecta el posicionamiento de nucleosomas, las muestras recogidas en cada etapa del ciclo celular fueron enviadas a secuenciar y se generaron los mapas de nucleosomas. Cuando se compararon los patrones de nucleosomas, se observó que la organización nucleosomal es en general bastante resistente a

fluctuaciones a lo largo del ciclo celular, de acuerdo con estudios anteriores a baja resolución La disminución de la ocupación de nucleosomas adyacentes a la NFR es notable en las fases S y M, en comparación con las fases G1 y G2. Por otra parte, un análisis más detallado reveló que los nucleosomas -1 y 1 son más difusos en las fases S y M, mientras que la amplitud de las NFR no muestra un cambio aparente. El análisis a nivel de genoma de la dispersión de los nucleosomas revela que la fase S tiene un aumento global de la deslocalización. Por otro lado, en la fase G1 la cromatina muestra el mejor posicionamiento, tanto alrededor de los TSSs como a lo largo del genoma, seguido por lo observado en la fase G2.

En general, hemos demostrado que la cromatina tiene una organización distinta en la fase S, particularmente en un determinado período de tiempo, que se traduce en una mayor sensibilidad a la MNasa y por tanto en una mayor imprecisión de los nucleosomas a lo largo del genoma, lo que podría indicar una mayor accesibilidad de la cromatina en este punto particular del ciclo en levadura.

El análisis de transición de los perfiles de nucleosomas entre etapas del ciclo celular demuestra que casi la mitad de los genes tienden a mantener un buen posicionamiento de los nucleosomas -1 y 1. Sin embargo, los genes que cambian su patrón nucleosomal alrededor del TSS tienen generalmente nucleosomas bien posicionados en las fases G1 y G2 que se alteran en las fases S y M. Adicionalmente, los genes con mayores cambios en el TSS en la fase G1 están involucrados en la respuesta a feromonas, la conjugación y la reproducción sexual. Por otro lado, alrededor del 70% de los genes preservan el estado de la NFR, entre los cuales encontramos un 39% de genes con las NFR abiertas contra un 32% de NFR cerradas. Los genes con la NFR cerrada son altamente enriquecidos en genes con una caja TATA, mientras que en genes con la NFR abierta el comportamiento es opuesto, acorde con estudios previos (Tirosh & Barkai 2008; Zaugg & Luscombe 2011). Curiosamente, los genes con NFR abiertas tienden a estar bajo el control del remodelador de la cromatina INO80 y ISW1.

La comparación de los perfiles de nucleosomas entre genes alta y bajamente expresados reveló claras diferencias. Con tasas de transcripción elevadas, la ocupación nucleosomal disminuye alrededor de regiones promotoras. Respecto a las regiones codificantes, los nucleosomas se encuentran más deslocalizados en genes altamente expresados. Esto

también es cierto para los nucleosomas aguas arriba (*upstream* en inglés) del promotor de los genes menos expresados. Por otra parte, se observó un aumento de la cobertura de nucleosomas en regiones codificantes con niveles de expresión más altos, notablemente en las fases S y M. Finalmente, hemos ampliado nuestro estudio mediante el examen de genes individuales en base a su expresión y arquitectura nucleosomal. En particular, se examinaron los genes que muestran una arquitectura nucleosomal diferencial así como una expresión variable de sus niveles de mRNA a lo largo del ciclo celular. Hemos encontrado una buena concordancia entre el estado de la transcripción y el patrón de nucleosomas en algunos genes dependientes del ciclo celular.

4. Organización de la cromatina en torno a los orígenes de replicación y centrómeros

El patrón nucleosomal alrededor de los orígenes de replicación de levadura se ha estudiado en gran detalle (Berbenetz et al. 2010; Eaton et al. 2010; Hoggard et al. 2013). A pesar de que la unión de proteínas replicadoras y su activación son procesos dinámicos, pocos estudios se han realizado en cuanto a la dinámica de la cromatina en torno a los orígenes de replicación en base al ciclo celular y además, la mayoría de los patrones de nucleosomas de origen reportados proceden de células asíncronas o en G1. Por esta razón, hemos examinado y comparado los mapas de posicionamiento de los nucleosomas alrededor de los orígenes de replicación en cada fase del ciclo celular. Los mapas se generaron por triplicado a partir de cultivos sincronizadas mediante la alineación de 253 orígenes anotados en relación con sitios de unión a complejos de origen de replicación (ORC por sus siglas en inglés) o a secuencias consenso de replicación autónoma (ACS del inglés) (Eaton et al. 2010). Como era de esperar, el patrón nucleosomal de los orígenes de replicación es bastante similar al de los TSSs a lo largo del ciclo celular, excepto que las NFRs en ACS están flanqueadas por nucleosomas bien posicionados de una manera simétrica, a diferencia del perfil en TSSs, y que las NFRs alrededor de los orígenes son más anchas que en las 5 'NFRs. Por otra parte, a pesar de que el perfil en general es bastante similar a lo largo del ciclo celular, hay ligeras variaciones en distintas fases particulares. La ocupación nucleosomal es más baja en las fases G1 y S en comparación con las fases G2 y M. Por otra parte, el análisis de los perfiles de orígenes individuales muestra que los orígenes en fase G1 contienen un

mayor número de nucleosomas en fase y una NFR con una mayor anchura promedio.

Los perfiles de orígenes individuales también demuestran la diversidad del patrón nucleosomal, tal como apuntan otros estudios (Berbenetz et al. 2010). Sobre todo, es notable la discrepancia en la anchura de las NFRs, que ocupa entre 80 y 600 pb. Puesto que se propuso que las diferencias en el tiempo de activación de los orígenes podrían ser el resultado de diferencias en la estructura de la cromatina, analizamos la relación entre el perfil nucleosomal y el periodo de activación del origen. Aunque no hemos podido demostrar ninguna relación entre la anchura de las NFR y el momento de origen de replicación, la eliminación gradual y la ocupación de los nucleosomas que flanquean las NFRs de los orígenes es mayor en los tempranos. Además, demostramos que los orígenes tempranos contienen nucleosomas con niveles más altos de la variante de histona H2A.Z.

Por otra parte, también estamos interesados en la estructura nucleosomal centromérica, ya que la composición exacta y la estructura del nucleosoma centromérico no se ha resuelto por completo todavía. Muchos estudios mediante técnicas diversas han propuesto modelos contrapuestos para explicar la estructura de los centrómeros. Entre la gran variedad de modelos, tres modelos sobresalen en la configuración nucleosomal centromérica: i) octámero, donde el nucleosoma octamérico contiene Cse4 en lugar de H3; ii) hexasoma, que implica la proteína SCM3 junto con el tetrámero Cse4/H4; iii) y hemisoma, que contiene una sola molécula de cada histona: Cse4 , H4, H2A y H2B La razón por la que la estructura del centrómero es todavía motivo de discusión se debe en parte a las dificultades técnicas. En la mayoría de los estudios, la histona Cse4 se sobre-expresa con etiquetas en el C-terminal o bien se basan en reconstituciones de nucleosomas in vitro, lo que podría sesgar la estructura real del nucleosoma. Por esa razón, hemos tratado de utilizar un nuevo enfoque para llevar a cabo un estudio de los nucleosomas centroméricos en su estado fisiológico.

Para estudiar la estructura centromérica en detalle, primero hemos analizado la señal nucleosomal centromérica. Cabe destacar que la señal centromérica en todos los conjuntos de datos se centra siempre en el punto medio de la secuencia centromérica (punto "0"), de acuerdo con su localización dependiente de secuencia. Por otra parte, es notable que el pico

nucleosomal centromérico es más estrecho que los picos nucleosomales circundantes. Mientras que el nucleosoma centromérico es de alrededor 100 pb, los nucleosomas que lo flanquean tienen 120-140 pb de ancho. De todos modos, los perfiles de nucleosomas centroméricos se obtuvieron por un protocolo de asignación de nucleosoma convencional. Y puesto que los nucleosomas centroméricos son menos estables que los canónicos, como se señaló anteriormente, algunos de los nucleosomas centroméricos podrían haberse disociado durante la generación de los mapas nucleosomales. Por lo tanto, para evitar estos problemas, se modificó el protocolo de preparación de nucleosomas, añadiendo una centrifugación en gradiente de sacarosa para separar los nucleosomas centroméricos de los canónicos en base a su diferencia de densidad. Primero intentamos enriquecer los nucleosomas que contienen cromatina centromérica utilizando anticuerpos contra H3 N-terminal (H3-N) y recoger el material no unido, ya que el N-terminal de Cse4 no comparte similitud con el H3-N. Luego cargamos el material no unido en un gradiente 5-15% de sacarosa. La presencia de nucleosoma centromérico se inspeccionó por PCR en las fracciones recogidas y el contenido de proteínas se analizó por tinción de plata en gel de SDS-PAGE. En consecuencia, los fragmentos que contienen nucleosomas centroméricos tienen una banda extra de alrededor 25 kDa, que se puede corresponder con Cse4 y demostrar así que el contenido de histonas es diferente. Además, carecen de la banda H3, lo que confirma la sustitución de H3 por Cse4. Sin embargo, las fracciones contienen bandas para H2B, H2A y H4 (aunque en este caso sea débil). Estos hallazgos demuestran que la composición nucleosomal centromérica es distinta de los nucleosomas canónicos y descartan pues modelo hexasoma, que implica dos moléculas de SCM3 junto con el tetrámero Cse4/H4.

Discusión

El objetivo de esta tesis es la comprensión de los mecanismos que determinan el posicionamiento de nucleosomas en el genoma, tales como la contribución de los factores que actúan *en cis*, la dinámica de los nucleosomas a lo largo del ciclo celular y su papel regulador en la expresión génica.

En primer lugar, se ha analizado el posicionamiento de nucleosomas en el genoma de *S. cerevisiae*. Para ello, hemos generado mapas nucleosomales a

nivel genómico basados en dos metodologías: la secuenciación masiva (MNase-Seq) y microarrays en mosaico (MNase-CHIP). Nuestro análisis comparativo pone de manifiesto que la mayoría de las regiones o bien empobrecidas o bien ricas en nucleosomas, están conservadas, pero hay algo de disimilitud entre los mapas. Dado que la tecnología MNase-Seq proporciona una mayor resolución, carece de la dependencia de la sonda, permite un análisis más preciso de las regiones no repetitivas pertinentes y una cobertura de casi todo el genoma, basamos la mayoría de nuestro estudio sobre mapas nucleosomas derivados de MNase-Seq.

De acuerdo con observaciones anteriores, hemos demostrado que los mapas nucleosomales muestran un posicionamiento no uniforme a lo largo del genoma de la levadura y exhiben una arquitectura nucleosomal canónica alrededor de los sitios de inicio de transcripción (TSSs), que consiste en un NFR 5' flanqueado por dos nucleosomas bien posicionados. Por otro lado, los perfiles de nucleosomas alrededor de los sitios de terminación de transcripción (TTSs) están peor definidos, donde regiones aguas abajo (*downstream* en inglés) del TTS, también conocidas como 3'NFRs, están empobrecidas en nucleosomas, mientras que *upstream* de los TTSs y *downstream* de las 3'NFRs están muy enriquecidas.

A continuación exploramos el papel de las propiedades físicas del ADN en el posicionamiento de los nucleosomas en torno a los TSSs y TTSs, como un posible factor determinante no aleatorio. Nuestro análisis de los perfiles de digestión por MNase de ADN genómico y nucleosomal reveló que los sitios de corte de la MNasa exhiben una composición tetramérica de nucleótidos muy similares. Por otra parte, la MNase degrada regiones denominadas LRs, las cuales son bastante comunes en el ADN genómico y nucleosomal y se definen como regiones altamente flexibles flanqueadas por motivos rígidos. En conjunto, nuestro análisis del perfil de digestión de MNasa a nivel del genoma demuestra que regiones reguladoras clave como las 5 ' y 3' NFRS están en realidad señalizadas mediante un patrón diferencial de susceptibilidad de MNasa, indicando que propiedades físicas intrínsecas del ADN determinan las principales localizaciones de nucleosomas en estas regiones. Sin embargo, debido a otros factores que juegan un papel en la regulación de genes y el posicionamiento de nucleosomas, las propiedades físicas del ADN sólo pueden conferir un mecanismo basal de la regulación génica.

Además, a pesar de que el patrón de nucleosoma canónico se conserve entre los diferentes mapas, hay diferencias e incluso entre genes de un mismo mapa. Por lo tanto, el desarrollo de modelos predictivos para un mapa nucleosomal consenso y la exploración de los factores determinantes del posicionamiento de nucleosomas son tareas complejas. Tratamos de minimizar los factores extrínsecos derivados de las condiciones experimentales, de procesamiento de datos o de la heterogeneidad de las muestras, que pueden introducir variación entre los mapas de nucleosomas. En primer lugar, exploramos una fuente importante de variación, los niveles de digestión con MNasa, debido a la alta actividad de la enzima y su dependencia de secuencia. Nuestros análisis indican que el exceso de digestión con MNasa conduce a la escisión dentro del nucleosoma, mientras que mapas poco digeridos presentan fragmentos más largos que pueden interferir con la alineación, lo que lleva a un mayor ruido, especialmente en la región de enlace entre nucleosomas. Investigamos entonces las variaciones potenciales que pueden surgir de la heterogeneidad celular mediante la comparación de perfiles de nucleosomas entre cultivos sincronizados y no sincronizados, observando un aumento general de deslocalización en los nucleosomas -1 y +1, especialmente para aquellos genes dependientes del ciclo celular. Esto indica que la asincronía puede producir una cantidad considerable de ruido debido a las variaciones de la expresión génica dependiente del ciclo celular, la compactación de la cromatina o la relajación a lo largo del ciclo celular. En base a nuestros resultados, reducimos al mínimo el ruido debido a la asincronía y a los niveles de digestión por MNasa para explorar las variaciones intrínsecas en el posicionamiento de nucleosomas, comparando mapas nucleosomales derivados de dos réplicas sincronizadas en fase G1 tardía. Se encontró que sólo el 3% de los nucleosomas se encuentran exactamente en la misma posición a lo largo del genoma, señalando su dinámica intrínseca. Por lo tanto, los nucleosomas bien posicionados pueden no estar bien fijados, sino que son móviles a lo largo del genoma, ya que las barreras intrínsecas deslizantes son muy bajas. Por último, hemos estudiado los determinantes del posicionamiento de nucleosomas con mayor precisión a partir del conjunto de genes seleccionados derivados de las muestras sincronizadas en fase G1 tardía. De acuerdo con ello, se demuestra que las propiedades físicas del ADN y los factores *trans* actúan de forma sinérgica para definir la NFR y por lo tanto, las matrices de nucleosomas. Por otro lado, lejos de las 5 'NFRs,

los nucleosomas son altamente dinámicos, cambiando su ubicación constantemente debido a la ausencia de una barrera energética.

En la segunda parte de nuestro estudio, el objetivo consiste en analizar la dinámica de la cromatina dependiente del ciclo celular. Por esa razón, hemos generado mapas de nucleosomas derivados de cultivos sincronizados en cada fase del ciclo celular. En primer lugar, comparamos mapas de nucleosomas con distintos patrones de digestión MNase en cada fase del ciclo celular, lo que revela que los nucleosomas en fase S están más desubicados a lo largo de genoma y tienen una mayor sensibilidad a la digestión con MNasa, en comparación con las fases G1, G2 y M. También se muestra que la condensación / descondensación de la cromatina a lo largo de ciclo celular se lleva a cabo de forma secuencial, ya que hay un pico de sensibilidad máxima en un punto concreto de la fase S. Estos resultados muestran que la fase S posee una organización de la cromatina distinta, con nucleosomas difusos y una mayor sensibilidad a la MNasa, y que estos cambios conformacionales son repentinos y agudos. Por otro lado, la cromatina en fase G1 muestra un buen posicionamiento alrededor de los TSSs y a lo largo del genoma, seguido por la fase G2. Puesto que las muestras en G1 están detenidas con el factor alfa, presentan el más alto nivel de sincronía, que explicaría en parte el mayor número de nucleosomas bien posicionados. El análisis de la transición de perfiles de nucleosomas entre las etapas del ciclo celular demuestra que casi la mitad de los genes mantienen el posicionamiento de los nucleosomas -1 y 1 a lo largo del ciclo celular, lo que indica que sólo ciertos genes se ven afectados por la dinámica de la cromatina en base al ciclo celular. Cuando se comparan las variaciones globales de cobertura nucleosomal entre etapas, observamos que la mayor variación en la cobertura de los nucleosomas en base al ciclo celular se produce en las transiciones de G1 a S y de G2 a M, mientras que la transición entre las fases S y G2 muestra mucha menos variación. Las variaciones en la cobertura nucleosomal tienen la misma tendencia que las variaciones de expresión entre etapas, destacando la interacción entre la cromatina y la transcripción.

Por otra parte, los perfiles de nucleosomas alrededor de los orígenes de replicación nos muestran que están muy conservados a lo largo del ciclo celular. Sin embargo, la anchura de las NFRs es un poco más grande en fase G1, donde se forma el complejo pre-replicativo (pre-RC en inglés) y por lo

tanto, protege a fragmentos de ADN más largos. No hemos podido demostrar ninguna relación entre el patrón de nucleosomas y el momento de origen de replicación, aunque los orígenes tempranos tienden a tener los nucleosomas mejor ubicados.

Por último, en base a nuestros ensayos de caracterización de nucleosomas centroméricos en condiciones fisiológicas, hemos sido capaces de demostrar que la composición nucleosomal centromérica es distinta a la de los nucleosomas canónicos y descartan el modelo hexasoma, que involucra a dos moléculas de SCM3 junto con el tetrámero Cse4/H4. Además, la composición de los nucleosomas centroméricos podría fluctuar a lo largo del ciclo celular.

REFERENCES

References

Albert, I. et al., 2007. Translational and rotational settings of H2A. Z nucleosomes across the Saccharomyces cerevisiae genome. *Nature*, 446(7135), pp.572–576.

Alexander, M., Heppel, L.A. & Hurwitz, J., 1961. The purification and properties of micrococcal nuclease. *Journal of Biological Chemistry*, 236(11), pp.3014–3019.

Allshire, R.C. & Karpen, G.H., 2008. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nature reviews. Genetics*, 9(12), pp.923–37.

Althoefer, H. et al., 1995. Mcm1 is required to coordinate G2-specific transcription in Saccharomyces cerevisiae. *Molecular and cellular biology*, 15(11), pp.5917–28.

Altman, R. & Kellogg, D., 1997. Control of mitotic events by Nap1 and the Gin4 kinase. *The Journal of cell biology*, 138(1), pp.119–130.

Anderson, J.D. & Widom, J., 2001. Poly (dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Molecular and cellular biology*, 21(11), pp.3830–3839.

Andrews, B. & Measday, V., 1998. The cyclin family of budding yeast: abundant use of a good idea. *Trends in Genetics*, 14(2), pp.66–72.

Aparicio, O.M., Stout, A.M. & Bell, S.P., 1999. Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. *Proceedings of the National Academy of Sciences*, 96(16), pp.9130–9135.

Aparicio, O.M., Weinstein, D.M. & Bell, S.P., 1997. Components and dynamics of DNA replication complexes in S. cerevisiae: redistribution of MCM proteins and Cdc45p during S phase. *Cell*, 91(1), pp.59–69.

Aravamudhan, P., Felzer-Kim, I. & Joglekar, A.P., 2013. The budding yeast point centromere associates with two Cse4 molecules during mitosis. *Current biology : CB*, 23(9), pp.770–4.

Badis, G. et al., 2008. A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Molecular Cell*, 32(6), pp.878–887.

Bai, L. & Morozov, A. V, 2010. Gene regulation by nucleosome positioning. *Trends in Genetics*, 26(11), pp.476–483.

Van Bakel, H. et al., 2013. A compendium of nucleosome and transcript profiles reveals determinants of chromatin architecture and transcription. *PLoS genetics*, 9(5), p.e1003479.

Bao, Y., White, C.L. & Luger, K., 2006. Nucleosome core particles containing a poly (dA· dT) sequence element exhibit a locally distorted DNA structure. *Journal of Molecular Biology*, 361(4), pp.617–624.

Bardin, A.J. & Amon, A., 2001. Men and sin: what's the difference? *Nature reviews. Molecular cell biology*, 2(11), pp.815–26.

Basehoar, A.D., Zanton, S.J. & Pugh, B.F., 2004. Identification and distinct regulation of yeast TATA box-containing genes. *Cell*, 116(5), pp.699–709.

Battistini, F. et al., 2010. Structural mechanics of DNA wrapping in the nucleosome. *Journal of Molecular Biology*, 396(2), pp.264–279.

Battistini, F. et al., 2012. Structure-Based Identification of New High-Affinity Nucleosome Binding Sequences. *Journal of Molecular Biology*, 420(1), pp.8–16.

Bazett-Jones, D.P., 1992. Electron spectroscopic imaging of chromatic and other nucleoprotein complexes. *Electron microscopy reviews*, 5(1), pp.37–58.

Bednar, J. et al., 1998. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proceedings of the National Academy of Sciences*, 95(24), pp.14173–14178.

Belch, Y. et al., 2010. Weakly positioned nucleosomes enhance the transcriptional competency of chromatin. *PLoS One*, 5(9), p.e12984.

Bell, S.P. & Stillman, B., 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature*, 357(6374), pp.128–34.

Berbenetz, N.M., Nislow, C. & Brown, G.W., 2010. Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure. *PLoS genetics*, 6(9), p.e1001092.

Bintu, L. et al., 2011. The elongation rate of RNA polymerase determines the fate of transcribed nucleosomes. *Nature Structural & Molecular Biology*, 18(12), pp.1394–1399.

Black, B.E. et al., 2004. Structural determinants for generating centromeric chromatin. *Nature*, 430(6999), pp.578–582.

Black, B.E. & Cleveland, D.W., 2011. Epigenetic centromere propagation and the nature of CENP-A nucleosomes. *Cell*, 144(4), pp.471–479.

Bloom, K.S. & Carbon, J., 1982. Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes. *Cell*, 29(2), pp.305–317.

Bock, L.J. et al., 2012. Cnn1 inhibits the interactions between the KMN complexes of the yeast kinetochore. *Nature Cell Biology*, 14(6), pp.614–624.

Bode, J. et al., 2003. From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. *Chromosome Research*, 11(5), pp.435–445.

Boeger, H. et al., 2003. Nucleosomes unfold completely at a transcriptionally active promoter. *Molecular Cell*, 11(6), pp.1587–1598.

Boeger, H., Griesenbeck, J. & Kornberg, R.D., 2008. Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. *Cell*, 133(4), pp.716–726.

Breeden, L., 1996. Start-specific transcription in yeast. In *Transcriptional Control of Cell Growth*. Springer, pp. 95–127.

Brewer, B.J., Chlebowicz-Sledziewska, E. & Fangman, W.L., 1984. Cell cycle phases in the unequal mother/daughter cell cycles of Saccharomyces cerevisiae. *Molecular and cellular biology*, 4(11), pp.2529–31.

De Bruin, R.A.. et al., 2004. Cln3 Activates G1-Specific Transcription via Phosphorylation of the SBF Bound Repressor Whi5. *Cell*, 117(7), pp.887–898.

Bryne, J.C. et al., 2008. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic acids research*, 36(Database issue), pp.D102–6.

Bui, M. et al., 2012. Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo. *Cell*, 150(2), pp.317–326.

Bui, M. et al., 2013. The CENP-A nucleosome: A battle between Dr Jekyll and Mr Hyde. *Nucleus*, 4(1), pp.10–11.

Cadoret, J.-C. et al., 2008. Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proceedings of the National Academy of Sciences*, 105(41), pp.15837–15842.

Cai, M. & Davis, R.W., 1990. Yeast centromere binding protein CBF1, of the helix-loop-helix protein family, is required for chromosome stability and methionine prototrophy. *Cell*, 61(3), pp.437–446.

Cairns, B.R., 2009. The logic of chromatin architecture and remodelling at promoters. *Nature*, 461(7261), pp.193–198.

Camahort, R. et al., 2009. Cse4 is part of an octameric nucleosome in budding yeast. *Molecular Cell*, 35(6), pp.794–805.

Carbon, J. & Clarke, L., 1984. Structural and functional analysis of a yeast centromere (CEN3). *Journal of Cell Science*, 1984(Supplement 1), pp.43–58.

Caterino, T.L. & Hayes, J.J., 2010. Structure of the H1 C-terminal domain and function in chromatin condensationThis paper is one of a selection of papers published in a Special Issue entitled 31st Annual International Asilomar Chromatin and Chromosomes Conference, and has undergone the Jou. *Biochemistry and Cell Biology*, 89(1), pp.35–44.

Cheeseman, I.M. & Desai, A., 2008. Molecular architecture of the kinetochore–microtubule interface. *Nature Reviews Molecular Cell Biology*, 9(1), pp.33–46.

Cho, U.-S. & Harrison, S.C., 2011. Ndc10 is a platform for inner kinetochore assembly in budding yeast. *Nature Structural & Molecular Biology*, 19(1), pp.48–55.

Choi, J.K. & Kim, Y.-J., 2008. Epigenetic regulation and the variability of gene expression. *Nature genetics*, 40(2), pp.141–147.

Chung, H.-R. & Vingron, M., 2009. Sequence-dependent nucleosome positioning. *Journal of Molecular Biology*, 386(5), pp.1411–1422.

Clarke, L. & Carbon, J., 1980. Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature*, 287(5782), pp.504–9.

Clayton, A.L., Hazzalin, C.A. & Mahadevan, L.C., 2006. Enhanced histone acetylation and transcription: a dynamic perspective. *Molecular Cell*, 23(3), pp.289–296.

Cleveland, D.W., Mao, Y. & Sullivan, K.F., 2003. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell*, 112(4), pp.407–421.

Cole, H.A., Howard, B.H. & Clark, D.J., 2011. The centromeric nucleosome of budding yeast is perfectly positioned and covers the entire centromere. *Proceedings of the National Academy of Sciences*, 108(31), pp.12687–12692.

Collins, K.A. et al., 2005. De novo kinetochore assembly requires the centromeric histone H3 variant. *Molecular biology of the cell*, 16(12), pp.5649–5660.

Crotti, L.B. & Basrai, M.A., 2004. Functional roles for evolutionarily conserved Spt4p at centromeres and heterochromatin in Saccharomyces cerevisiae. *The EMBO journal*, 23(8), pp.1804–1814.

Dabney, J. & Meyer, M., 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques*, 52(2), pp.87–94.

Dalal, Y. et al., 2007. Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. *PLoS biology*, 5(8), p.e218.

Davey, C.A. et al., 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *Journal of Molecular Biology*, 319(5), pp.1097–1113.

Davie, J.R. et al., 1981. Histone modifications in the yeast S. cerevisiae. *Nucleic acids research*, 9(13), pp.3205–3216.

Dechassa, M.L. et al., 2011. Structure and Scm3-mediated assembly of budding yeast centromeric nucleosomes. *Nature communications*, 2, p.313.

Dekker, J., 2008. Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. *Journal of Biological Chemistry*, 283(50), pp.34532–34540.

Dimitriadis, E.K. et al., 2010. Tetrameric organization of vertebrate centromeric nucleosomes. *Proceedings of the National Academy of Sciences*, 107(47), pp.20317–20322.

Dingwall, C., Lomonossoff, G.P. & Laskey, R.A., 1981. High sequence specificity of micrococcal nuclease. *Nucleic acids research*, 9(12), pp.2659–2674.

Dorigo, B. et al., 2004. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science*, 306(5701), pp.1571–1573.

Douglas Maya, M.M.-H.S.C. and M.-C.M.-C., 2011. Free Histones and the Cell Cycle. In H. Seligmann, ed. *DNA Replication-Current Advances*. InTech.

Drew, H.R. & Travers, A.A., 1985. DNA bending and its relation to nucleosome positioning. *Journal of Molecular Biology*, 186(4), pp.773–790.

Dunleavy, E., Pidoux, A. & Allshire, R., 2005. Centromeric chromatin makes its mark. *Trends in biochemical sciences*, 30(4), pp.172–175.

Eaton, M.L. et al., 2010. Conserved nucleosome positioning defines replication origins. *Genes & development*, 24(8), pp.748–753.

Espelin, C.W., Kaplan, K.B. & Sorger, P.K., 1997. Probing the architecture of a simple kinetochore using DNA–protein crosslinking. *The Journal of cell biology*, 139(6), pp.1383–1396.

Evans, T. et al., 1983. Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell*, 33(2), pp.389–396.

Felsenfeld, G., 1992. Chromatin as an essential part of the transcriptional mechanim. *Nature*, 355, pp.219–224.

Ferguson, B.M. et al., 1991. A yeast origin of replication is activated late in S phase. *Cell*, 65(3), pp.507–515.

Fillingham, J. et al., 2008. Chaperone control of the activity and specificity of the histone H3 acetyltransferase Rtt109. *Molecular and cellular biology*, 28(13), pp.4342–4353.

Finch, J.T. & Klug, A., 1976. Solenoidal model for superstructure in chromatin. *Proceedings of the National Academy of Sciences*, 73(6), pp.1897–1901.

Fisher, A.G. & Merkenschlager, M., 2002. Gene silencing, cell fate and nuclear organisation. *Current opinion in genetics & development*, 12(2), pp.193–197.

Fitch, I. et al., 1992. Characterization of four B-type cyclin genes of the budding yeast Saccharomyces cerevisiae. *Molecular biology of the cell*, 3(7), p.805.

Fitzgerald-Hayes, M., Clarke, L. & Carbon, J., 1982. Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. *Cell*, 29(1), pp.235–244.

Flanagan, J.F. & Peterson, C.L., 1999. A role for the yeast SWI/SNF complex in DNA replication. *Nucleic acids research*, 27(9), pp.2022–2028.

Flemming, W., 1882. Zellsubstanz. *Kern und Zellteilung*.

Flick, J.T., Eissenberg, J.C. & Elgin, S.C.R., 1986. Micrococcal nuclease as a DNA structural probe: its recognition sequences, their genomic distribution and correlation with DNA structure determinants. *Journal of Molecular Biology*, 190(4), pp.619–633.

Flores, O. & Orozco, M., 2011. nucleR: a package for non-parametric nucleosome positioning. *Bioinformatics*, 27(15), pp.2149–2150.

Foltz, D.R. et al., 2006. The human CENP-A centromeric nucleosome-associated complex. *Nature Cell Biology*, 8(5), pp.458–469.

Fragoso, G. et al., 1995. Nucleosome positioning on the MMTV LTR results from the frequency-biased occupancy of multiple frames. *Genes & development*, 9(15), pp.1933–1947.

Friedman, K.L. et al., 1996. Multiple determinants controlling activation of yeast replication origins late in S phase. *Genes & development*, 10(13), pp.1595–1607.

Furuyama, S. & Biggins, S., 2007. Centromere identity is specified by a single centromeric nucleosome in budding yeast. *Proceedings of the National Academy of Sciences*, 104(37), pp.14706–14711.

Furuyama, T. & Henikoff, S., 2009. Centromeric nucleosomes induce positive DNA supercoils. *Cell*, 138(1), pp.104–113.

Fussner, E., Ching, R.W. & Bazett-Jones, D.P., 2011. Living without 30nm chromatin fibers. *Trends in biochemical sciences*, 36(1), pp.1–6.

Gascoigne, K.E. & Cheeseman, I.M., 2012. T time for point centromeres. *Nature Cell Biology*, 14(6), pp.559–561.

Gauthier, N.P. et al., 2010. Cyclebase.org: version 2.0, an updated comprehensive, multi-species repository of cell cycle experiments and derived analysis results. *Nucleic acids research*, 38(Database issue), pp.D699–702.

Gentleman, R.C. et al., 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5(10), p.R80.

Georgieva, M. et al., 2012. Hho1p, the linker histone of Saccharomyces cerevisiae, is important for the proper chromatin organization in vivo. *Biochimica et biophysica acta*, 1819(5), p.366.

Gerard, R. & Gluzman, Y., 1986. Functional analysis of the role of the A + T-rich region and upstream flanking sequences in simian virus 40 DNA replication. *Molecular and cellular biology*, 6(12), pp.4570–7.

Gkikopoulos, T. et al., 2011. The SWI/SNF complex acts to constrain distribution of the centromeric histone variant Cse4. *The EMBO journal*, 30(10), pp.1919–1927.

Glowczewski, L. et al., 2000. Histone-histone interactions and centromere function. *Molecular and cellular biology*, 20(15), pp.5700–5711.

Goñi, J.R. et al., 2007. Determining promoter location based on DNA structure first-principles calculations. *Genome biology*, 8(12), p.R263.

Goñi, J.R. et al., 2008. DNAlive: a tool for the physical analysis of DNA at the genomic scale. *Bioinformatics*, 24(15), pp.1731–1732.

Gregan, J. et al., 2003. Fission yeast Cdc23/Mcm10 functions after pre-replicative complex formation to promote Cdc45 chromatin binding. *Molecular biology of the cell*, 14(9), pp.3876–3887.

Grigoryev, S.A., Bednar, J. & Woodcock, C.L., 1999. MENT, a heterochromatin protein that mediates higher order chromatin folding, is a new serpin family member. *The Journal of biological chemistry*, 274(9), pp.5626–36.

Gruss, C. et al., 1993. Disruption of the nucleosomes at the replication fork. *The EMBO journal*, 12(12), p.4533.

Guillemette, B. et al., 2005. Variant histone H2A. Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS biology*, 3(12), p.e384.

Guillemette, B. & Gaudreau, L., 2006. Reuniting the contrasting functions of H2A.Z. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 84(4), pp.528–35.

Han, M. & Grunstein, M., 1988. Nucleosome loss activates yeast downstream promoters in vivo. *Cell*, 55(6), pp.1137–1145.

Hancock, R., 2000. A new look at the nuclear matrix. *Chromosoma*, 109(4), pp.219–225.

Harp, J.M. et al., 2000. Asymmetries in the nucleosome core particle at 2.5 A resolution. *Acta Crystallographica Section D: Biological Crystallography*, 56(12), pp.1513–1534.

Hartley, P.D. & Madhani, H.D., 2009. Mechanisms that specify promoter nucleosome location and identity. *Cell*, 137(3), pp.445–458.

Hartwell, L.H., Culotti, J. & Reid, B., 1970. Genetic control of the cell-division cycle in yeast, I. Detection of mutants. *Proceedings of the National Academy of Sciences*, 66(2), pp.352–359.

Hartwell, L.H. & Unger, M.W., 1977. Unequal division in Saccharomyces cerevisiae and its implications for the control of cell division. *The Journal of cell biology*, 75(2 Pt 1), pp.422–35.

Hassan, A.H., Neely, K.E. & Workman, J.L., 2001. Histone acetyltransferase complexes stabilize swi/snf binding to promoter nucleosomes. *Cell*, 104(6), pp.817–827.

Henikoff, S. et al., 2000. Heterochromatic deposition of centromeric histone H3-like proteins. *Proceedings of the National Academy of Sciences*, 97(2), pp.716–721.

Henikoff, S. & Dalal, Y., 2005. Centromeric chromatin: what makes it unique? *Current opinion in genetics & development*, 15(2), pp.177–184.

Hogan, G.J., Lee, C.-K. & Lieb, J.D., 2006. Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters M. Snyder, ed. *PLoS genetics*, 2(9), p.e158.

Hoggard, T. et al., 2013. A Link between ORC-Origin Binding Mechanisms and Origin Activation Time Revealed in Budding Yeast. *PLoS genetics*, 9(9), p.e1003798.

Van Holde, K.E., Chromatin. *Springer series in molecular biology*.

Horowitz, R.A. et al., 1994. The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. *The Journal of cell biology*, 125(1), pp.1–10.

Hörz, W. & Altenburger, W., 1981. Sequence specific cleavage of DNA by micrococcal nuclease. *Nucleic acids research*, 9(12), pp.2643–2658.

Huang, S. et al., 2005. Rtt106p is a histone chaperone involved in heterochromatin-mediated silencing. *Proceedings of the National*

Academy of Sciences of the United States of America, 102(38), pp.13410–13415.

Huebert, D.J. et al., 2012. Dynamic changes in nucleosome occupancy are not predictive of gene expression dynamics but are linked to transcription and chromatin regulators. *Molecular and cellular biology*, 32(9), pp.1645–1653.

Hughes, A. & Rando, O.J., 2009. Chromatin "programming" by sequence--is there more to the nucleosome code than %GC? *Journal of biology*, 8(11), p.96.

Hughes, A.L. et al., 2012. A Functional Evolutionary Approach to Identify Determinants of Nucleosome Positioning: A Unifying Model for Establishing the Genome-wide Pattern. *Molecular Cell*, 48(1), pp.5–15.

Ioshikhes, I.P. et al., 2006. Nucleosome positions predicted through comparative genomics. *Nature genetics*, 38(10), pp.1210–1215.

Irniger, S., 2002. Cyclin destruction in mitosis: a crucial task of Cdc20. *FEBS letters*, 532(1), pp.7–11.

Ishii, K., 2009. Conservation and divergence of centromere specification in yeast. *Current opinion in microbiology*, 12(6), pp.616–622.

Jansen, A. & Verstrepen, K.J., 2011. Nucleosome positioning in Saccharomyces cerevisiae. *Microbiology and Molecular Biology Reviews*, 75(2), pp.301–320.

Jiang, C. & Pugh, B.F., 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nature Reviews Genetics*, 10(3), pp.161–172.

Joti, Y. et al., 2012. Chromosomes without a 30-nm chromatin fiber. *Nucleus*, 3(5), pp.404–410.

Kaplan, N. et al., 2008. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, 458(7236), pp.362–366.

Kaplan, N. et al., 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, 458(7236), pp.362–6.

Keith, K.C. & Fitzgerald-Hayes, M., 2000. CSE4 genetically interacts with the Saccharomyces cerevisiae centromere DNA elements CDE I and CDE II but not CDE III: implications for the path of the centromere DNA around a Cse4p variant nucleosome. *Genetics*, 156(3), pp.973–981.

Khorasanizadeh, S., 2004. The nucleosome: from genomic organization to genomic regulation. *Cell*, 116(2), pp.259–272.

Kingston, I.J., Yung, J.S.Y. & Singleton, M.R., 2011. Biophysical characterization of the centromere-specific nucleosome from budding yeast. *Journal of Biological Chemistry*, 286(5), pp.4021–4026.

Kizilyaprak, C. et al., 2010. In vivo chromatin organization of mouse rod photoreceptors correlates with histone modifications. *PLoS One*, 5(6), p.e11039.

Knott, S.R. V et al., 2012. Forkhead transcription factors establish origin timing and long-range clustering in S. cerevisiae. *Cell*, 148(1-2), p.99.

Knott, S.R. V et al., 2009. Genome-wide replication profiles indicate an expansive role for Rpd3L in regulating replication initiation timing or efficiency, and reveal genomic loci of Rpd3 function in Saccharomyces cerevisiae. *Genes & development*, 23(9), pp.1077–1090.

Korber, P. et al., 2004. Evidence for histone eviction in trans upon induction of the yeast PHO5 promoter. *Molecular and cellular biology*, 24(24), pp.10965–10974.

Kornberg, R.D., 1974. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science*, 184(4139), pp.868–871.

Kornberg, R.D. & Lorch, Y., 1999. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*, 98(3), pp.285–94.

Kornberg, R.D. & Stryer, L., 1988. Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic acids research*, 16(14), pp.6677–6690.

Krassovsky, K., Henikoff, J.G. & Henikoff, S., 2012. Tripartite organization of centromeric chromatin in budding yeast. *Proceedings of the National Academy of Sciences*, 109(1), pp.243–248.

Krude, T., 1999. Chromatin assembly during DNA replication in somatic cells. *European Journal of Biochemistry*, 263(1), pp.1–5.

Krude, T., 1995. Chromatin assembly factor 1 (CAF-1) colocalizes with replication foci in HeLa cell nuclei. *Experimental cell research*, 220(2), pp.304–311.

Kruithof, M. et al., 2009. Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nature Structural & Molecular Biology*, 16(5), pp.534–540.

Kuan, P.F. et al., 2009. A non-homogeneous hidden-state model on first order differences for automatic detection of nucleosome positions. *Statistical applications in genetics and molecular biology*, 8(1), pp.1–45.

Lampert, F. & Westermann, S., 2011. A blueprint for kinetochores—new insights into the molecular mechanics of cell division. *Nature Reviews Molecular Cell Biology*, 12(7), pp.407–412.

Lavery, R. et al., 2009. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic acids research*, 37(17), pp.5917–5929.

Lechner, J., 1994. A zinc finger protein, essential for chromosome segregation, constitutes a putative DNA binding subunit of the Saccharomyces cerevisiae kinetochore complex, Cbf3. *The EMBO journal*, 13(21), p.5203.

Lechner, J. & Carbon, J., 1991. A 240 kd multisubunit protein complex, CBF3, is a major component of the budding yeast centromere. *Cell*, 64(4), pp.717–725.

Lee, C.-K. et al., 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature genetics*, 36(8), pp.900–905.

Lee, W. et al., 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nature genetics*, 39(10), pp.1235–1244.

Lefrançois, P. et al., 2009. Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing. *BMC genomics*, 10(1), p.37.

Lehner, B., 2010. Conflict between noise and plasticity in yeast. *PLoS genetics*, 6(11), p.e1001185.

Li, F. et al., 2004. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14), pp.4781–4786.

Lipford, J.R. & Bell, S.P., 2001. Nucleosomes positioned by ORC facilitate the initiation of DNA replication. *Molecular Cell*, 7(1), pp.21–30.

Lõoke, M. et al., 2013. Chromatin-dependent and -independent regulation of DNA replication origin activation in budding yeast. *EMBO reports*, 14(2), pp.191–8.

Lowary, P.T. & Widom, J., 1989. Higher-order structure of Saccharomyces cerevisiae chromatin. *Proceedings of the National Academy of Sciences*, 86(21), pp.8266–8270.

Lowary, P.T. & Widom, J., 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology*, 276(1), pp.19–42.

Luger, K. et al., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), pp.251–260.

Luger, K., Dechassa, M.L. & Tremethick, D.J., 2012. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature Reviews Molecular Cell Biology*, 13(7), pp.436–447.

Luger, K. & Richmond, T.J., 1998. DNA binding within the nucleosome core. *Current opinion in structural biology*, 8(1), pp.33–40.

MacAlpine, H.K. et al., 2010. Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome research*, 20(2), pp.201–211.

Malvezzi, F. et al., 2013. A structural basis for kinetochore recruitment of the Ndc80 complex via two distinct centromere receptors. *The EMBO journal*, 32(3), pp.409–23.

Masumoto, H. et al., 2005. A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature*, 436(7048), pp.294–8.

Mavrich, T.N. et al., 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome research*, 18(7), pp.1073–1083.

McGREW, J., Diehl, B. & Fitzgerald-Hayes, M., 1986. Single base-pair mutations in centromere element III cause aberrant chromosome segregation in Saccharomyces cerevisiae. *Molecular and cellular biology*, 6(2), pp.530–538.

Mellor, J. & Morillon, A., 2004. ISWI complexes in Saccharomyces cerevisiae. *Biochimica et biophysica acta*, 1677(1-3), p.100.

Meluh, P.B. et al., 1998. Cse4p Is a Component of the Core Centromere of< i> Saccharomyces cerevisiae</i>. *Cell*, 94(5), pp.607–613.

Miele, V. et al., 2008. DNA physical properties determine nucleosome occupancy from yeast to fly. *Nucleic acids research*, 36(11), pp.3746–3756.

Mizuguchi, G. et al., 2007. Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes. *Cell*, 129(6), pp.1153–1164.

Morgan, D.O., 2007. *The cell cycle: principles of control*, New Science Press.

Morozov, A. V et al., 2009. Using DNA mechanics to predict in vitro nucleosome positions and formation energies. *Nucleic acids research*, 37(14), pp.4707–4722.

Morse, R.H., 1989. Nucleosomes inhibit both transcriptional initiation and elongation by RNA polymerase III in vitro. *The EMBO journal*, 8(8), p.2343.

Murray, A.W. & Hunt, T., 1993. *The cell cycle: an introduction*, Oxford University Press New York/Oxford.

Muthurajan, U.M. et al., 2004. Crystal structures of histone Sin mutant nucleosomes reveal altered protein-DNA interactions. *The EMBO journal*, 23(2), pp.260–71.

Nasmyth, K., Peters, J.-M. & Uhlmann, F., 2000. Splitting the chromosome: cutting the ties that bind sister chromatids. *Science*, 288(5470), pp.1379–1384.

Nieduszynski, C.A. et al., 2007. OriDB: a DNA replication origin database. *Nucleic acids research*, 35(suppl 1), pp.D40–D46.

Nishino, Y. et al., 2012. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO journal*, 31(7), pp.1644–53.

Nurse, P., Thuriaux, P. & Nasmyth, K., 1976. Genetic control of the cell division cycle in the fission yeast Schizosaccharomyces pombe. *Molecular and General Genetics MGG*, 146(2), pp.167–178.

Olins, A.L. & Olins, D.E., 1974. Spheroid chromatin units (ν bodies). *Science*, 183(4122), pp.330–332.

Olson, W.K. et al., 1998. DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proceedings of the National Academy of Sciences*, 95(19), pp.11163–11168.

Ong, M.S., Richmond, T.J. & Davey, C.A., 2007. DNA stretching and extreme kinking in the nucleosome core. *Journal of Molecular Biology*, 368(4), pp.1067–1074.

Palmer, D.K. et al., 1987. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *The Journal of cell biology*, 104(4), pp.805–815.

Papamichos-Chronakis, M. & Peterson, C.L., 2008. The Ino80 chromatin-remodeling enzyme regulates replisome function and stability. *Nature Structural & Molecular Biology*, 15(4), pp.338–345.

Pappas, D.L., Frisch, R. & Weinreich, M., 2004. The NAD+-dependent Sir2p histone deacetylase is a negative regulator of chromosomal DNA replication. *Genes & development*, 18(7), pp.769–781.

Park, Y.-J. et al., 2005. Nucleosome assembly protein 1 exchanges histone H2A-H2B dimers and assists nucleosome sliding. *Journal of Biological Chemistry*, 280(3), pp.1817–1825.

Parnell, T.J., Huff, J.T. & Cairns, B.R., 2007. RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. *The EMBO journal*, 27(1), pp.100–110.

Pearson, C.G. et al., 2004. Stable kinetochore-microtubule attachment constrains centromere positioning in metaphase. *Current Biology*, 14(21), pp.1962–1967.

Peckham, H.E. et al., 2007. Nucleosome positioning signals in genomic DNA. *Genome research*, 17(8), pp.1170–1177.

Pereira, G. et al., 2002. Regulation of the Bfa1p-Bub2p complex at spindle pole bodies by the cell cycle phosphatase Cdc14p. *The Journal of cell biology*, 157(3), pp.367–79.

Peterson, C.L. & Logie, C., 2000. Recruitment of chromatin remodeling machines. *Journal of cellular biochemistry*, 78(2), pp.179–185.

Pietrasanta, L.I. et al., 1999. Probing the Saccharomyces cerevisiae centromeric DNA (CEN DNA)–binding factor 3 (CBF3) kinetochore complex by using atomic force microscopy. *Proceedings of the National Academy of Sciences*, 96(7), pp.3757–3762.

Planet, E. et al., 2012. htSeqTools: high-throughput sequencing quality control, processing and visualization in R. *Bioinformatics*, 28(4), pp.589–590.

Pluta, A.F. et al., 1995. The centromere: hub of chromosomal activities. *Science*, 270(5242), pp.1591–1594.

Pokholok, D.K. et al., 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*, 122(4), pp.517–527.

Polach, K.J., Lowary, P.T. & Widom, J., 2000. Effects of core histone tail domains on the equilibrium constants for dynamic DNA site accessibility in nucleosomes. *Journal of Molecular Biology*, 298(2), pp.211–223.

Purvis, A. & Singleton, M.R., 2008. Insights into kinetochore-DNA interactions from the structure of Cep3Delta. *EMBO reports*, 9(1), pp.56–62.

Radman-Livaja, M. & Rando, O.J., 2010. Nucleosome positioning: how is it established, and why does it matter? *Developmental biology*, 339(2), pp.258–266.

Raghuraman, M.K. et al., 2001. Replication dynamics of the yeast genome. *Science*, 294(5540), pp.115–121.

Raisner, R.M. et al., 2005. Histone variant H2A. Z marks the 5 ends of both active and inactive genes in euchromatin. *Cell*, 123(2), pp.233–248.

Rando, O.J. & Ahmad, K., 2007. Rules and regulation in the primary structure of chromatin. *Current opinion in cell biology*, 19(3), pp.250–256.

Rando, O.J. & Chang, H.Y., 2009. Genome-wide views of chromatin structure. *Annual review of biochemistry*, 78, p.245.

Rando, O.J. & Winston, F., 2012. Chromatin and transcription in yeast. *Genetics*, 190(2), pp.351–87.

Ransom, M., Dennehey, B.K. & Tyler, J.K., 2010. Chaperoning histones during DNA replication and repair. *Cell*, 140(2), pp.183–195.

Richmond, T.J. & Davey, C.A., 2003. The structure of DNA in the nucleosome core. *Nature*, 423(6936), pp.145–150.

Robinson, P.J.J. et al., 2006. EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proceedings of the National Academy of Sciences*, 103(17), pp.6506–6511.

Rohs, R. et al., 2010. Origins of specificity in protein-DNA recognition. *Annual review of biochemistry*, 79, pp.233–69.

Da Rosa, J.L. et al., 2011. Overlapping regulation of CenH3 localization and histone H3 turnover by CAF-1 and HIR proteins in Saccharomyces cerevisiae. *Genetics*, 187(1), pp.9–19.

Rudner, A.D. & Murray, A.W., 2000. Phosphorylation by Cdc28 activates the Cdc20-dependent activity of the anaphase-promoting complex. *The Journal of cell biology*, 149(7), pp.1377–1390.

Sadeh, R. & Allis, C.D., 2011. Genome-wide "re"-modeling of nucleosome positions. *Cell*, 147(2), pp.263–6.

Sakabe, K. & Okazaki, R., 1966. A unique property of the replicating region of chromosomal DNA. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis*, 129(3), pp.651–654.

Schalch, T. et al., 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 436(7047), pp.138–141.

Scheffer, M.P., Eltsov, M. & Frangakis, A.S., 2011. Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei.

Proceedings of the National Academy of Sciences, 108(41), pp.16992–16997.

Schones, D.E. et al., 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell*, 132(5), pp.887–898.

Segal, E. et al., 2006. A genomic code for nucleosome positioning. *Nature*, 442(7104), pp.772–778.

Segal, E. & Widom, J., 2009. Poly (dA: dT) tracts: major determinants of nucleosome organization. *Current opinion in structural biology*, 19(1), pp.65–71.

Sekinger, E.A., Moqtaderi, Z. & Struhl, K., 2005. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Molecular Cell*, 18(6), pp.735–748.

Sekulic, N. et al., 2010. The structure of (CENP-A-H4) 2 reveals physical features that mark centromeres. *Nature*, 467(7313), pp.347–351.

Shimada, K. et al., 2008. Ino80 chromatin remodeling complex promotes recovery of stalled replication forks. *Current Biology*, 18(8), pp.566–575.

Shivaraju, M. et al., 2012. Cell-cycle-coupled structural oscillation of centromeric nucleosomes in yeast. *Cell*, 150(2), pp.304–316.

Shivaswamy, S. et al., 2008. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS biology*, 6(3), p.e65.

Simpson, R.T., 1990. Nucleosome positioning can affect the function of a cis-acting DNA element in vivo. *Nature*, 343(6256), pp.387–9.

Smith, S. & Stillman, B., 1989. Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. *Cell*, 58(1), pp.15–25.

Spellman, P.T. et al., 1998. Comprehensive identification of cell cycle–regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Molecular biology of the cell*, 9(12), pp.3273–3297.

Stegmeier, F., Visintin, R. & Amon, A., 2002. Separase, polo kinase, the kinetochore protein Slk19, and Spo12 function in a network that controls Cdc14 localization during early anaphase. *Cell*, 108(2), pp.207–220.

Stemmann, O. & Lechner, J., 1996. The Saccharomyces cerevisiae kinetochore contains a cyclin-CDK complexing homologue, as identified by in vitro reconstitution. *The EMBO journal*, 15(14), p.3611.

Stocking, C.R. & Gifford, E.M., 1959. Incorporation of thymidine into chloroplasts of Spirogyra. *Biochemical and Biophysical Research Communications*, 1(3), pp.159–164.

Stoler, S. et al., 2007. Scm3, an essential Saccharomyces cerevisiae centromere protein required for G2/M progression and Cse4 localization. *Proceedings of the National Academy of Sciences*, 104(25), pp.10571–10576.

Struhl, K. & Segal, E., 2013. Determinants of nucleosome positioning. *Nature structural & molecular biology*, 20(3), pp.267–73.

Sullivan, K.F., Hechenberger, M. & Masri, K., 1994. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *The Journal of cell biology*, 127(3), pp.581–592.

Suto, Robert K et al., 2000. Crystal structure of a nucleosome core particle containing the variant histone H2A. Z. *Nature Structural & Molecular Biology*, 7(12), pp.1121–1124.

Suto, R K et al., 2000. Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nature structural biology*, 7(12), pp.1121–4.

Szerlong, H.J. & Hansen, J.C., 2010. Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structureThis paper is one of a selection of papers published in a Special Issue entitled 31st Annual International Asilomar Chromatin and Chromosomes Conference, and. *Biochemistry and Cell Biology*, 89(1), pp.24–34.

Takahashi, K., Chen, E.S. & Yanagida, M., 2000. Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. *Science*, 288(5474), pp.2215–2219.

Tanaka, S., Livingstone-Zatchej, M. & Thoma, F., 1996. Chromatin structure of the yeast URA3 gene at high resolution provides insight into structure and positioning of nucleosomes in the chromosomal context. *Journal of Molecular Biology*, 257(5), pp.919–934.

Thåström, A., Lowary, P.T. & Widom, J., 2004. Measurement of histone–DNA interaction free energy in nucleosomes. *Methods*, 33(1), pp.33–44.

Thoma, F., Bergman, L.W. & Simpson, R.T., 1984. Nuclease digestion of circular TRP1ARS1 chromatin reveals positioned nucleosomes separated by nuclease-sensitive regions. *Journal of Molecular Biology*, 177(4), pp.715–733.

Tirosh, I. & Barkai, N., 2008. Two strategies for gene regulation by promoter nucleosomes. *Genome research*, 18(7), pp.1084–1091.

Tirosh, I., Sigal, N. & Barkai, N., 2010. Widespread remodeling of mid-coding sequence nucleosomes by Isw1. *Genome biology*, 11(5), p.R49.

Tjong, H. et al., 2012. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome research*, 22(7), pp.1295–1305.

Tolkunov, D. & Morozov, A. V, 2010. Genomic studies and computational predictions of nucleosome positions and formation energies. *Advances in Protein Chemistry and Structural Biology*, 79, pp.1–57.

Tolstorukov, M.Y. et al., 2007. A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. *Journal of Molecular Biology*, 371(3), pp.725–738.

Trifonov, E.N., 2010. Base pair stacking in nucleosome DNA and bendability sequence pattern. *Journal of theoretical biology*, 263(3), pp.337–339.

Trifonov, E.N. & Sussman, J.L., 1980. The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proceedings of the National Academy of Sciences of the United States of America*, 77(7), pp.3816–20.

Tsankov, A.M. et al., 2010. The role of nucleosome positioning in the evolution of gene regulation. *PLoS biology*, 8(7), p.e1000414.

Tsunaka, Y. et al., 2005. Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. *Nucleic acids research*, 33(10), pp.3424–3434.

Tyler, J.K. et al., 1999. The RCAF complex mediates chromatin assembly during DNA replication and repair. *Nature*, 402(6761), pp.555–560.

Unnikrishnan, A., Gafken, P.R. & Tsukiyama, T., 2010. Dynamic changes in histone acetylation regulate origins of DNA replication. *Nature Structural & Molecular Biology*, 17(4), pp.430–437.

Vaillant, C., Audit, B. & Arneodo, A., 2007. Experiments confirm the influence of genome long-range correlations on nucleosome positioning. *Physical review letters*, 99(21), p.218103.

Venters, B.J. & Pugh, B.F., 2009. A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome research*, 19(3), pp.360–371.

Verdaasdonk, J.S. & Bloom, K., 2011. Centromeres: unique chromatin structures that drive chromosome segregation. *Nature Reviews Molecular Cell Biology*, 12(5), pp.320–332.

Verreault, A., 2000. De novo nucleosome assembly: new pieces in an old puzzle. *Genes & development*, 14(12), pp.1430–1438.

Vincent, J.A., Kwong, T.J. & Tsukiyama, T., 2008. ATP-dependent chromatin remodeling shapes the DNA replication landscape. *Nature Structural & Molecular Biology*, 15(5), pp.477–484.

Visintin, R. et al., 1998. The phosphatase Cdc14 triggers mitotic exit by reversal of Cdk-dependent phosphorylation. *Molecular Cell*, 2(6), pp.709–718.

Vogelauer, M. et al., 2002. Histone acetylation regulates the time of replication origin firing. *Molecular Cell*, 10(5), pp.1223–1233.

Vujcic, M., Miller, C.A. & Kowalski, D., 1999. Activation of silent replication origins at autonomously replicating sequence elements near the HML locus in budding yeast. *Molecular and cellular biology*, 19(9), pp.6098–6109.

Warburton, P.E. et al., 1997. Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. *Current Biology*, 7(11), pp.901–904.

Weiner, A. et al., 2010. High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome research*, 20(1), pp.90–100.

Weinreich, M., Palacios DeBeer, M.A. & Fox, C.A., 2004. The activities of eukaryotic replication origins in chromatin. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1677(1), pp.142–157.

White, C.L., Suto, R.K. & Luger, K., 2001. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *The EMBO journal*, 20(18), pp.5207–5218.

Whitehouse, I. et al., 2003. Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. *Molecular and cellular biology*, 23(6), pp.1935–1945.

Whitehouse, I. et al., 1999. Nucleosome mobilization catalysed by the yeast SWI/SNF complex. *Nature*, 400(6746), pp.784–787.

Whitehouse, I. & Tsukiyama, T., 2006. Antagonistic forces that position nucleosomes in vivo. *Nature Structural & Molecular Biology*, 13(7), pp.633–640.

Widom, J. & Klug, A., 1985. Structure of the 300A chromatin filament: X-ray diffraction from oriented samples. *Cell*, 43(1), pp.207–13.

Williams, S.P. et al., 1986. Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. *Biophysical journal*, 49(1), pp.233–248.

Wolffe, A., 1998. *Chromatin: structure and function*, Access Online via Elsevier.

Woodcock, C.L. & Ghosh, R.P., 2010. Chromatin higher-order structure and dynamics. *Cold Spring Harbor perspectives in biology*, 2(5), p.a000596.

Wyrick, J.J. et al., 1999. Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. *Nature*, 402(6760), pp.418–421.

Wyrick, J.J. et al., 2001. Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. *Science*, 294(5550), pp.2357–2360.

Xiao, H. et al., 2011. Nonhistone Scm3 binds to AT-rich DNA to organize atypical centromeric nucleosome of budding yeast. *Molecular Cell*, 43(3), pp.369–380.

Xu, F. & Olson, W.K., 2010. DNA architecture, deformability, and nucleosome positioning. *Journal of biomolecular structure & dynamics*, 27(6), pp.725–39.

Yabuki, N., Terashima, H. & Kitada, K., 2002. Mapping of early firing origins on a replication profile of budding yeast. *Genes to Cells*, 7(8), pp.781–789.

Yen, K. et al., 2012. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell*, 149(7), pp.1461–1473.

Yin, S. et al., 2009. The impact of nucleosome positioning on the organization of replication origins in eukaryotes. *Biochemical and biophysical research communications*, 385(3), pp.363–368.

Yuan, G.-C. et al., 2005. Genome-scale identification of nucleosome positions in S. cerevisiae. *Science*, 309(5734), pp.626–630.

Zachariae, W., 1999. Progression into and out of mitosis. *Current opinion in cell biology*, 11(6), pp.708–716.

Zachariae, W. & Nasmyth, K., 1999. Whose end is destruction: cell division and the anaphase-promoting complex. *Genes & development*, 13(16), pp.2039–2058.

Zacher, B., Torkler, P. & Tresch, A., 2011. Analysis of Affymetrix ChIP-chip data using starr and R/Bioconductor. *Cold Spring Harbor protocols*, 2011(5), p.pdb.top110.

Zaugg, J.B. & Luscombe, N.M., 2012. A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast. *Genome research*, 22(1), pp.84–94.

Zaugg, J.B. & Luscombe, N.M., 2011. A genomic model of condition-specific nucleosome behaviour explains transcriptional activity in yeast. *Genome Research*, 22(1), p.gr.124099.111–.

Zawadzki, K.A., Morozov, A. V & Broach, J.R., 2009. Chromatin-dependent transcription factor accessibility rather than nucleosome remodeling predominates during global transcriptional restructuring in Saccharomyces cerevisiae. *Molecular biology of the cell*, 20(15), pp.3503–3513.

Zhang, H., Roberts, D.N. & Cairns, B.R., 2005. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell*, 123(2), pp.219–231.

Zhang, W., Colmenares, S.U. & Karpen, G.H., 2012. Assembly of Drosophila centromeric nucleosomes requires CID dimerization. *Molecular Cell*, 45(2), p.263.

Zhang, Y. et al., 2009. Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nature Structural & Molecular Biology*, 16(8), pp.847–852.

Zhang, Z. et al., 2011. A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science*, 332(6032), pp.977–980.

Zhang, Z., Shibahara, K. & Stillman, B., 2000. PCNA connects DNA replication to epigenetic inheritance in yeast. *Nature*, 408(6809), pp.221–225.

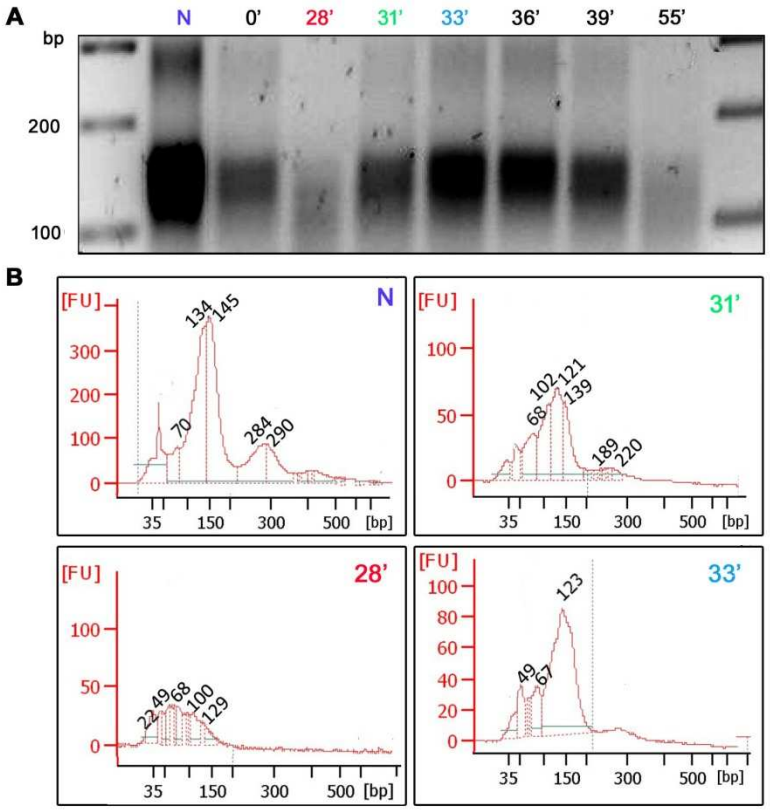
Zhou, J. et al., 2005. Cell cycle regulation of chromatin at an origin of DNA replication. *The EMBO journal*, 24(7), pp.1406–1417.

Zlatanova, J. et al., 2009. The nucleosome family: dynamic and growing. *Structure*, 17(2), pp.160–171.

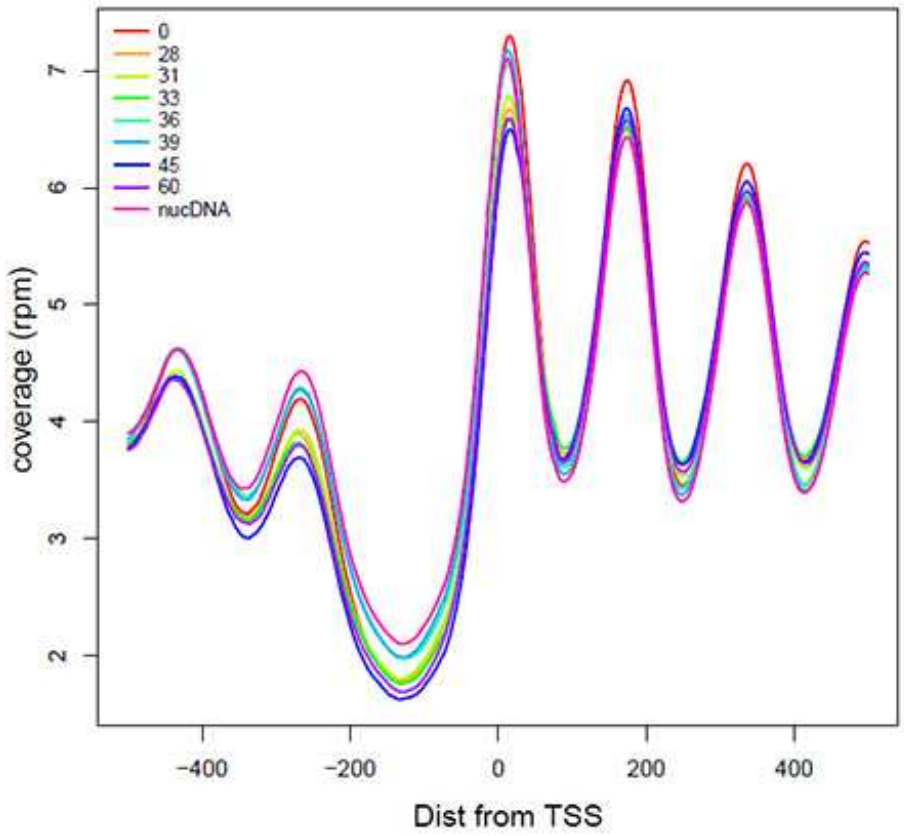
Zlatanova, J. & Thakar, A., 2008. H2A. Z: view from the top. *Structure*, 16(2), pp.166–179.

APPENDIX

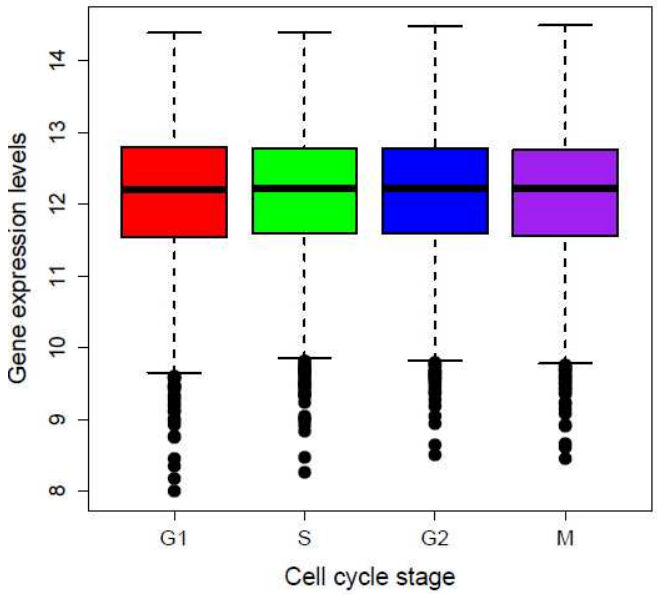
APPENDIX



Appendix 1: MNase cleavage assay along cell cycle at shorter intervals. A. Samples collected at indicated time points after alpha factor release and asynchronous sample are digested by 0.1 U MNase (25', 37°C) and fractionated on 2% agarose gel. B. Bioanalyzer profiles of the samples show their MNase digestion pattern.



Appendix 2: Comparison of nucleosome profiles along cell cycle at shorter intervals. Coverage maps per base pair were calculated and averaged across all yeast genome around TSSs at 0', 28', 31', 33', 36', 39', 45' and 60' after alpha factor release.



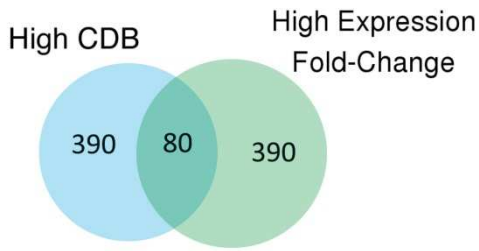
Appendix 3: Comparison of the average gene expression along cell cycle. Boxplot of the average expression for the 3279 genes at different cell cycle stages. Color boxes cover the 2nd and 3rd quartile, with mean value represented as a black line in the center. Gene expression levels are indicated as $\log_2$ values of the hybridization ratios from the Affymetrix GeneChip Yeast Genome 2.0 arrays.

Pvalue	Count	Term	class
1.5E-7	14	conjugation with cellular fusion	FWWW
1.4E-7	14	conjugation	FWWW
1.6E-7	13	response to pheromone	FWWW
1.8E-5	13	response to organic substance	FWWW
1.2E-4	15	sexual reproduction	FWWW
7.9E-3	7	response to pheromone during conjugation with cellular fusion	FWWW
2.1E-2	13	translational elongation	FWWW
8.3E-2	4	adaptation to pheromone during conjugation with cellular fusion	FWWW
1.1E-1	4	regulation of cell morphogenesis	FWWW

Appendix 4: Correlation of enriched GO terms with nucleosome fluctuation along cell cycle. The genes with fuzzy +1 nucleosomes at G1 and well-positioned nucleosomes at S, G2 and M phases are selected and analyzed for GO-terms enrichment.

Pvalue	Count	Term
8.8E-5	22	conjugation
4.3E-4	20	conjugation with cellular fusion
9.1E-4	18	response to pheromone
8.3E-4	31	sexual reproduction
2.5E-3	28	external encapsulating structure organization
2.5E-3	28	cell wall organization
4.6E-3	21	response to organic substance
3.4E-2	5	cell adhesion
3.4E-2	5	biological adhesion
6.2E-2	7	regulation of cell morphogenesis
9.0E-2	11	response to pheromone during conjugation with cellular fusion
2.8E-1	43	cell cycle

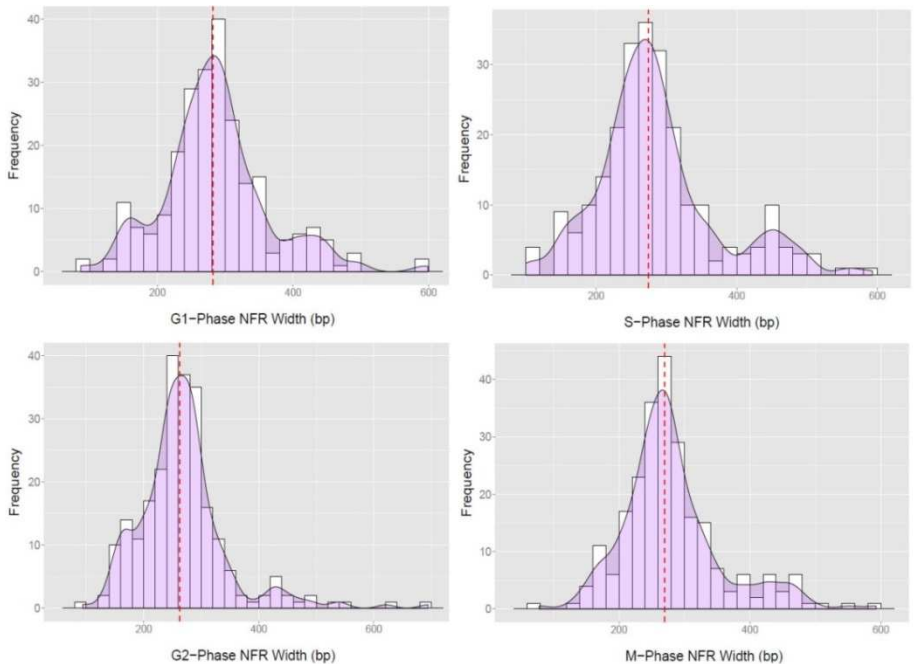
Appendix 5: Correlation of enriched GO terms with CDB between G1 and S phases. The top 5% of genes with highest CDB between G1 and S phases are selected and analyzed for GO-terms enrichment.



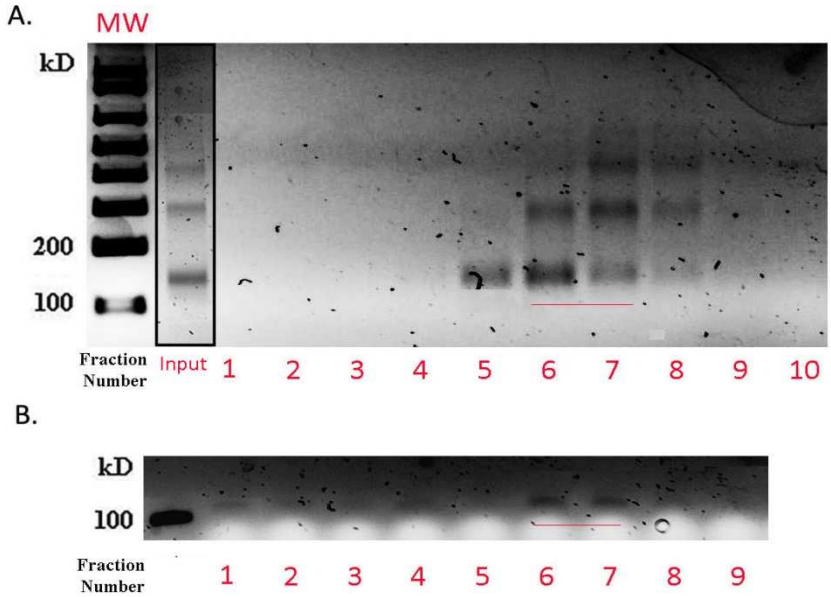
Appendix 6: Interplay between nucleosome plasticity and gene expression variation. The Venn diagram shows the intersection between the top high 10% nucleosome plasticity genes (measured as highest SD CDB) and the high expression-plastic genes, around TSS.

Pvalue	Count	Term
7.8E-9	16	response to pheromone
5.0E-8	16	conjugation
1.6E-7	17	response to organic substance
1.4E-7	15	conjugation with cellular fusion
5.8E-6	19	sexual reproduction
4.8E-5	10	response to pheromone during conjugation with cellular fusion
1.4E-2	13	reproductive cellular process
1.8E-1	4	karyogamy
2.2E-1	3	cytogamy
2.4E-1	5	cell surface receptor linked signal transduction
3.7E-1	8	regulation of cell cycle
4.5E-1	5	nucleus organization
5.6E-1	9	external encapsulating structure organization
5.9E-1	5	regulation of mitotic cell cycle

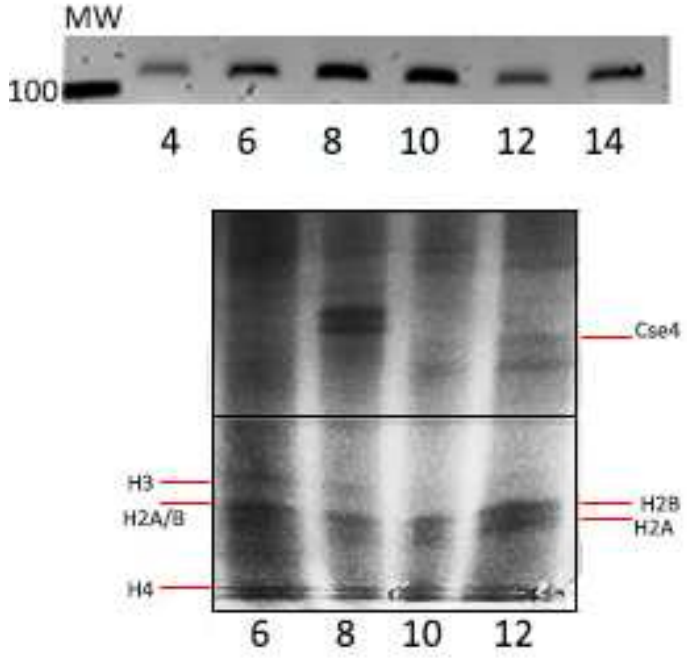
Appendix 7: Correlation of enriched GO terms with high expression and nucleosome plasticity. The top 10% of genes displaying high CDB and expression are selected and analyzed for GO-terms enrichment.



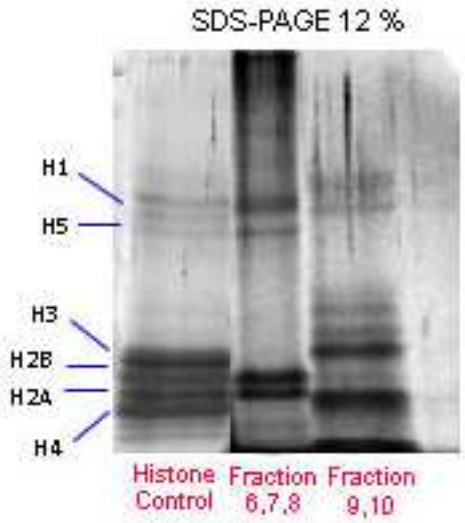
Appendix 8: NFR distribution around replication origins at G1, S, G2 and M phases. The plots demonstrate a similar distribution to normal distribution.



Appendix 9: Sucrose gradient (5-25%) fractionation of MNase digested chromatin. A. Portions of fractions were analyzed on 2% agarose gel. B. The presence of centromeric nucleosomes was determined by standard PCR using 5´CEN4 and 3´CEN4 and the PCR signal was analyzed on 2% agarose gel.



Appendix 10: Characterization of centromeric nucleosome. A. The centromeric nucleosome enrichment in fractions from sucrose gradient is determined by PCR. B. The protein content of each fraction is visualized by Silver Staining of 4-15% SDS-PAGE



Appendix 11: The protein content of each fraction is visualized by Silver Staining of 12% SDS-PAGE gel.