2014-11-14

Dissecting cis and trans Determinants of Nucleosome Positioning: A Dissertation

Amanda L. Hughes
University of Massachusetts Medical School

DISSECTING *CIS* AND *TRANS* DETERMINANTS OF NUCLEOSOME POSITIONING

A Dissertation Presented

By

AMANDA LAURA HUGHES

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

November 14th 2014

Biochemistry and Molecular Pharmacology

DISSECTING *CIS* AND *TRANS* DETERMINANTS OF NUCLEOSOME POSITIONING

A Dissertation Presented

By

AMANDA LAURA HUGHES

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

_____

Oliver J. Rando, M.D., Ph.D., Thesis Advisor

_____

Thomas Fazzio, Ph.D., Member of Committee

_____

Paul Kaufman, Ph.D., Member of Committee

_____

Melissa Moore, Ph.D., Member of Committee

_____

Zhiping Weng, Ph.D., Member of Committee

_____

Stephen Buratowski, Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

_____

Craig Peterson, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of the Biomedical Sciences signifies that the student has met all graduation requirements of the school

_____

Anthony Carruthers, Ph.D.,

Dean of the Graduate School of Biomedical Sciences

Program in Biochemistry and Molecular Pharmacology

November 14, 2014

ACKNOWLEDGEMENTS

I would like to acknowledge my thesis advisor, Oliver Rando, who always has his door open for discussion. I have always left his office feeling that I have a better understanding of where I want to go with my project. I thank the rest of the Rando lab members, past and present. The wonderful group of people, with whom I have been blessed to work, makes the Rando and Rhind labs a lively and pleasant working environment. In particular, I would like to thank Marta Radman-Livaja, who answered yeast and data analysis questions from the beginning, and Alex Tsankov, whose work formed the basis of a large part of my work. I would like to thank the members of my TRAC committee for suggestions and support throughout my time in graduate school. I am grateful to Tom Fazzio for joining my Dissertation Defense committee and look forward to his insights.

I would also like to thank the people with whom I have collaborated or who have shared data with me. Yi Jin's work made my studies possible, and our collaboration with the Struhl lab was a rewarding one. Data analysis from Assaf Weiner and Moran Yassour in the Friedman lab has provided the metric against which to measure my own later attempts. Towards the end of my research, I was given data from projects performed by Paolo Ferrari in the Strubin lab and Nils Krietenstein and Megha Wal in the Korber and Pugh labs. I very much enjoyed this opportunity to play with something new and different.

ABSTRACT

Eukaryotic DNA is packaged in chromatin, whose repeating subunit, the nucleosome, consists of an octamer of histone proteins wrapped by about 147bp of DNA. This packaging affects the accessibility of DNA and hence any process that occurs on DNA, such as replication, repair, and transcription. An early observation from genome-wide nucleosome mapping in yeast was that genes had a surprisingly characteristic structure, which has motivated studies to understand what determines this architecture. Both sequence and *trans* acting factors are known to influence chromatin packaging, but the relative contributions of *cis* and *trans* determinants of nucleosome positioning is debated. Here we present data using genetic approaches to examine the contributions of *cis* and *trans* acting factors on nucleosome positioning in budding yeast.

We developed the use of yeast artificial chromosomes to exploit quantitative differences in the chromatin structures of different yeast species. This allows us to place approximately 150kb of sequence from any species into the *S.cerevisiae* cellular environment and compare the nucleosome positions on this same sequence in different environments to discover what features are variant and hence regulated by *trans* acting factors. This method allowed us to conclusively show that the great preponderance of nucleosomes are positioned by *trans* acting factors. We observe the maintenance of nucleosome depletion over some promoter sequences, but partial fill-in of NDRs in some of the YAC

promoters indicates that even this feature is regulated to varying extents by *trans* acting factors.

We are able to extend our use of evolutionary divergence in order to search for specific *trans* regulators whose effects vary between the species. We find that a subset of transcription factors can compete with histones to help generate some NDRs, with clear effects documented in a *cbf1* deletion mutant. In addition, we find that Chd1p acts as a potential “molecular ruler” involved in defining the nucleosome repeat length differences between *S.cerevisiae* and *K.lactis*. The mechanism of this measurement is unclear as the alteration in activity is partially attributable to the N-terminal portion of the protein, for which there is no structural data. Our observations of a specialized chromatin structure at *de novo* transcriptional units along with results from nucleosome mapping in the absence of active transcription indicate that transcription plays a role in engineering genic nucleosome architecture. This work strongly supports the role of *trans* acting factors in setting up a dynamic, regulated chromatin structure that allows for robustness and fine-tuning of gene expression.

TABLE OF CONTENTS

Acknowledgments iii
Abstract iv
Table of Contents vi
List of Tables viii
List of Figures ix
List of Third Part Copyrighted Material xii
List of Abbreviations and Nomenclature xiii
Preface xiv
CHAPTER I: Introduction 1
Nucleosome Structure 1
How are nucleosome positions established *in* vivo? 17
Consequences of nucleosome positioning 45
CHAPTER II: High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. 59
Abstract 59
Introduction 60
Results 63
Discussion 104
Materials and Methods 109
CHAPTER III: A functional evolutionary approach to identify determinants of nucleosome positioning: A unifying model for establishing the genome-wide pattern. 112
Abstract 112
Introduction 113
Results 116
Discussion 145
Methods 152

CHAPTER IV: General Regulatory Factors play a role in promoter nucleosome depletion. 156
Introduction 156
Results 158
Discussion 164
Materials and Methods 166
CHAPTER V: Chd1p is a *trans*-acting factor that sets nucleosome spacing. 172
Introduction 172
Results 176
Discussion 190
Materials and Methods 192
CHAPTER VI: Discussion. 201
CHAPTER AI: Defining the order of the relationship between the locations of the TSS and +1 nucleosome. 211
Introduction 211
Results and Discussion 212
Materials and Methods 218
CHAPTER AII: The involvement of the preinitiation comples in promoter nucleosome occupancy and turnover. 226
Introduction 226
Results and Discussion 227
Materials and Methods 232
CHAPTER AIII: Mix n' matching whole cell extracts for *in vitro* reconstitution of chromatin structure. 236
Introduction 236
Results and Discussion 237
Materials and Methods 241
Bibliography 243

LIST OF TABLES

Table III.1 Species and YAC strain list.

Table IV.1 Primers used for Cbf1 strain construction.

Table V.1 Primers used for Chd1 deletion and “swap” strain construction.

Table V.2 Primers used to generate chimeric *CHD1.*

Table AI.1 Oligos used for constructing TSS mapping libraries.

LIST OF FIGURES

Figure I.1	Crystal structure of the nucleosome core particle.
Figure I.2	Canonical genic nucleosome architecture.
Figure I.3	Nucleosome positioning by sequence and *trans*-acting factors.
Figure I.4	ATP-dependent chromatin remodellers affect chromatin structure in several ways.
Figure I.5	Chromatin remodeler activity is targeted to different areas of the gene.
Figure I.6	Promoter chromatin organization allows for tuning of gene expression.
Figure II.1	Template Filtering overview
Supplementary Figure II.1	Templates used for filtering.
Supplementary Figure II.2	Nucleosome calls account for the majority of the sequencing data
Supplementary Figure II.3	Template filtering applied to other Illumina datasets
Supplementary Figure II.4	Comparison of template fitering and Parzen window-based approach
Supplementary Figure II.5	Differences between template filtering and Parzen window-based approach
Supplementary Figure II.6	Different templates capture MNase biases and delocalized nucleosomes
Figure II.2	Different promoter types are differently packaged
Supplementary Figure II.7	RNA polymerase accounts for some aspects of chromatin architecture
Figure II.3	Effects of MNase level on chromatin structure

Supplementary Figure II.8	Effects of MNase level on nucleosome occupancy and template usage
Supplementary Figure II.9	Polymerase inactivation results in globally inaccessible chromatin
Figure II.4	Effects of RNA polymerase on chromatin structure
Supplementary Figure II.10	Effects of RNA polymerase inactivation on gene category chromatin packing
Supplementary Figure II.11	Downstream nucleosome shifting is specific to Pol2 shutoff
Figure II.5	Nucleosomes relax towards *in vitro* preferred locations after Pol2 loss
Supplementary Figure II.12	Polymerase loss results in nucleosome shifts towards *in vitro* peaks
Supplementary Figure II.13	Polymerase loss eliminates short +1 to +3 spacing at ribosomal genes
Figure III.1	Functional evolutionary dissection of chromatin establishment mechanisms
Supplementary Figure III.1	NDRs are generally better-maintained over sequence from *K.lactis* than over *D.hansenii* sequence
Figure III.2	Promoter nucleosome depletion is maintained over poly(dA:dT) elements
Supplementary Figure III.2	Bulk chromatin is not affected in YAC-bearing strains
Figure III.3	Nucleosome spacing is set in *trans*
Figure III.4	+1 nucleosome shifts associated with transcriptional changes
Supplementary Figure III.3	+1 nucleosome shifts associated with transcription
Supplementary Figure III.4	Comparison of RNA-seq and MNase-seq datasets
Supplementary Figure III.5	Nucleosome positioning shifts are associated with shifts in TSS

Figure III.5	Characterisation of novel NDRs in YACs
Supplementary Figure III.6	Examples of new NDRs in *D.hansenii* YACs
Figure III.6	Three-step model for establishment of nucleosome positioning *in vivo*
Supplementary Figure III.7	Yeast whole cell extracts poorly position nucleosomes
Figure IV.1	Enrichment of transcription factor footprints in NDRs
Figure IV.2	Transcription-factor binding is associated with NDR formation
Figure IV.3	Cbf1 acts as a GRF in *D.hansenii*, but does not possess inherent nucleosome ordering properties
Figure V.1	*K.lactis* Chd1 generates increased internucleosome spacing throughout coding regions
Figure V.2	Chd1 spacing activity is greatest at the most highly-expressed genes.
Figure V.3	The N-terminus of Chd1 bears the majority of the differential measurement in the species
Figure V.4	N-terminus but not chromodomains is required for the increased measurement seen in the N-terminal chimera
Figure AI.1	The *sua7E62K* mutant does not shift TSS selection downstream genome-wide
Figure AI.2	Nucleosome positioning is affected in the *sua7E62K* mutant
Figure AII.1	TBP-depletion has a slight effect on promoter and genic nucleosome occupancy
Figure AII.2	Loss of TBP slightly decreases histone turnover
Figure AIII.1	Reconstitution with WCE from heterologous species does not generate *in vivo* nucleosome profiles

LIST OF THIRD PARTY COPYRIGHTED MATERIAL

Figure I.1	Nature Publishing Group, Macmillan Publishers Ltd…3487870483384
Figure I.3	Nature Publishing Group, Macmillan Publishers Ltd…3487870679949
Figure I.3	The American Association for the Advancement of Science…3487870776181
Figure I.5	Elsevier…3490870111186

LIST OF ABBREVIATIONS AND NOMENCLATURE

ATP	Adenosine triphosphate
CDS	coding region
ChIP	chromatin immunoprecipitation
GRF	General Regulatory Factor
MNase	Micrococcal Nuclease
NDR	Nucleosome Depleted Region
NFR	Nucleosome Free Region
PIC	preinitiation complex
TBP	TATA-binding protein
TF	Transcription Factor
TSS	Transcription Start Site
Turnover	replacement of a histone octamer with a newly synthesized (tagged) octamer
WCE	whole cell extract
YAC	yeast artificial chromosome

PREFACE

An abridged version of CHAPTER I has been published as:

Hughes, AL and Rando, OJ. (2014) Mechanisms underlying nucleosome positioning in vivo. Annu Rev Biophys. 43:41-63.

The work in CHAPTER II was a collaborative project between the Rando and Friedman labs, in which yeast growth, heat shock, and MNase-seq was performed by AL Hughes, and data analysis was developed and instrumented by A Weiner and M Yassour. This has been published as :

Weiner A, Hughes A, Yassour M, Rando OJ., and Friedman N. (2010) High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res. 20(1): 90-100.

The work in CHAPTER III was a collaboration between the Rando and Struhl labs. Y Jin generated the YAC strains and performed RNA-seq and TFIIB ChIP. AL Hughes performed MNase-seq, 5' RACE, and mapping of sequencing data. Y Jin and AL Hughes analysed data with guidance from K Struhl and OJ Rando. This has been published as:

Hughes AL*, Jin Y*, Rando OJ, Struhl K. (2012) A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. Mol Cell. 48(1):5-15.

The TSS mapping in CHAPTER AI was performed by Ting Ni in the Zhu lab, while MNase-seq and all data analysis was performed by AL Hughes.

The generation, growth, and H3/HA ChIP of TBP anchor away strains in CHAPTER AII was performed by Paolo Ferrari in the Strubin Lab. Subsequent generation of deep sequencing libraries and data analysis was performed by AL Hughes.

For CHAPTER AIII, chromatin reconstitution and nucleosomal DNA isolation was performed by Nils Krietenstein in the Korber lab, nucleosomal libraries were generated by Megha Wal in the Pugh lab, and data analysis was performed by AL Hughes.

Strain generation, MNase-seq, and data analysis in CHAPTER IV and CHAPTER V was all performed by AL Hughes, with the exception of Figure IV.1A, which was generated by Assaf Weiner.

CHAPTER I: Introduction

With the exception of dinoflagellates, all eukaryotes package their genomes into a repeating nucleoprotein complex known as the nucleosome. All eukaryotic DNA transactions, from transcription to DNA repair to recombination, occur in the context of this nucleosomal packaging. As the DNA wrapped around the nucleosome and the DNA located between nucleosomes differ in their accessibility and structural characteristics, the precise locations of nucleosomes have proven to be of great importance for understanding the function of the genome.

Nucleosome structure

The nucleosome core particle consists of ~147 base pairs of double-stranded DNA wrapped ~1.7 times in a left-handed superhelical turn around an octamer of histone proteins – 2 each of H2A, H2B, H3, and H4 (Kornberg and Lorch, 1999; Luger et al., 1997) (Figure I.1). The path of DNA around the histone octamer can vary between nucleosomes, as for example 145 bp templates can be wrapped in a nucleosome but require additional DNA stretching to accommodate the decreased total DNA incorporated (Ong et al., 2007).

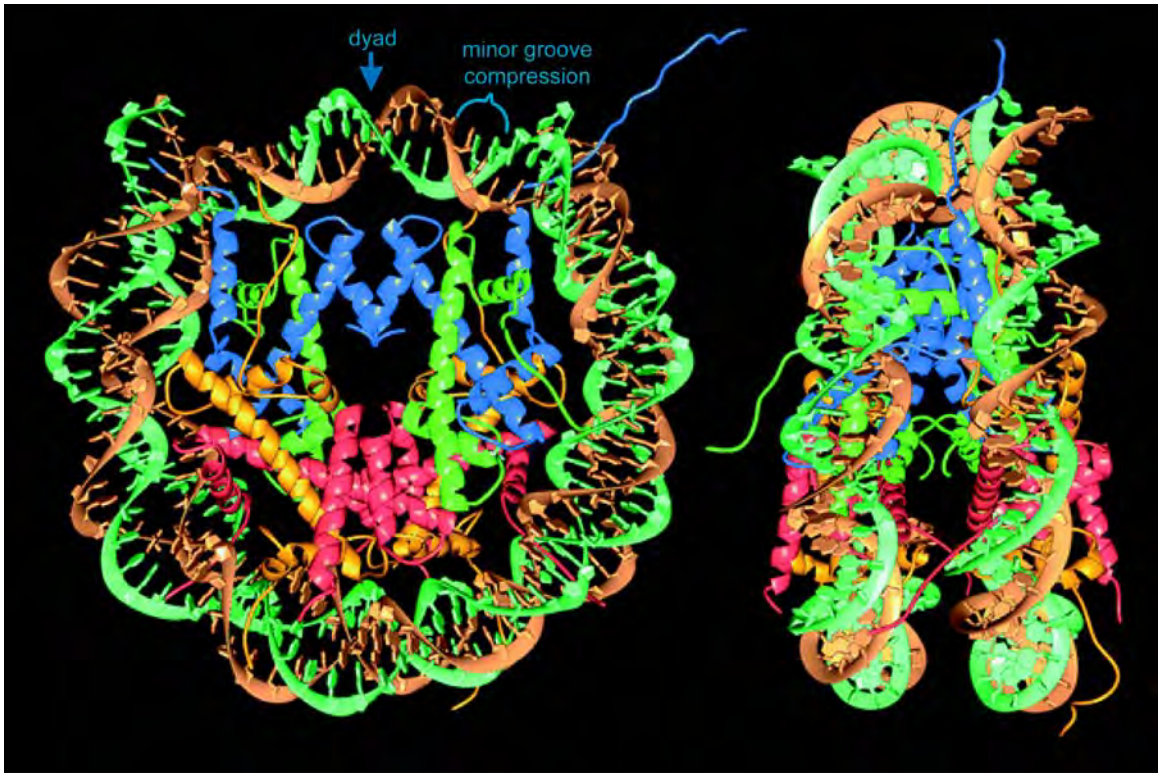


Figure I. 1: Crystal structure of the nucleosome core particle.

About 147bp of DNA wrap around an octamer of histone proteins—2 copies each of H2A (yellow), H2B (red), H3 (blue), and H4 (green). The H3/H4 dimers are joined to form a tetramer and mark the center or dyad of the nucleosome; H2A/H2B dimers attach via connection of H2A and H4 to form the octamer. The wrapping of double stranded DNA around the histones causes compression of the minor groove, when it is facing in towards the core.

The tight wrapping of DNA in nucleosomes has several important consequences. First of all, the superhelical path of DNA wrapped around the histones is significantly more curved than the persistence length of double-stranded DNA in solution. As a result of this deformation, intrinsically flexible DNA sequences are energetically favored for nucleosome formation, whereas stiff DNA sequences relatively disfavor nucleosomal incorporation. Second, the wrapping of DNA into nucleosomes has dramatic effects on the ability of DNA-binding proteins to access their template. On average, the "linker" DNA that lies between nucleosomes is far more accessible than the DNA in contact with the histones. In addition, DNA is not equally accessible throughout the nucleosome – DNA located near the entry/exit points is generally more accessible than DNA located near the dyad axis (Figure I.1). This accessibility is a result of thermal "breathing" of DNA at the edges of the nucleosome – detailed kinetic studies suggest that the DNA at the edges of the nucleosome binds and unbinds the histones over time scales of tens of milliseconds (Anderson and Widom, 2001; Anderson and Widom, 2000; Polach and Widom, 1995; Polach and Widom, 1996), and this breathing can have regulatory consequences. Importantly, the sequence of the DNA wrapped around the histone octamer, and the presence or absence of specific covalent histone modifications, can affect the rates of DNA breathing at the entry/exit points. Internally, the wrapping of DNA around the nucleosome results in alternating stretches in which the DNA major groove faces

either towards the octamer surface or outwards towards the solution, and this differential exposure can also be biologically relevant.

The many nucleosomes that package a given genome can differ from one another in many important ways. For example, in addition to the four “canonical” histone proteins – H2A, H2B, H3, and H4 – a number of variant histones can be assembled into nucleosomes. The number of variant histones found in the genome differs between species. Many smaller organisms such as budding yeast encode only two histone variants – a largely centromere-specific histone H3 variant (Cse4 in yeast, CENPA in mammals), and a predominantly promoter-localized H2A variant known as Htz1 or H2A.Z. Larger organisms often encode more histone variants, including relatively well-studied histone variants such as H3.3, MacroH2A, and H2A.X, as well as understudied variants such as H2A.Bbd and others. Many of these histone variants significantly alter the structure of the nucleosome. Most dramatically, it has been suggested that centromeric nucleosomes are composed of histone tetramers rather than octamers (Dalal et al., 2007), and are wrapped by a right-handed rather than a left-handed superhelix of DNA (Furuyama and Henikoff, 2009), although both of these fascinating hypotheses are the subject of ongoing and rather heated debates.

The specific subunit composition of the histone octamer can, in principle, influence any of the mechanisms involved in nucleosome positioning listed below. An increasing body of literature deals with the impact of histone variants

on ATP-dependent chromatin remodeling enzymes, but relatively little study has focused on the impact of histone variants on nucleosomal sequence preferences. Altogether, it is worth being aware that the classic nucleosome core particle crystal structure represents one family member among a multitude of related entities. In addition to histone variants, the exact DNA sequence wrapped around the histone octamer and the histone modifications present in the octamer can all alter key biophysical and structural characteristics of the nucleosome core particle and thereby exert regulatory effects on the genome.

In vivo nucleosome positions

The relative inaccessibility of nucleosomal DNA is the basis for several convenient experimental tools for identifying nucleosome positions, as linker DNA has proven to be far more nuclease-sensitive than nucleosomal DNA. Classically, this allowed researchers to link DNase-hypersensitive sites to gene regulatory elements (Stalder et al., 1980; Weintraub and Groudine, 1976) which are strongly nucleosome-depleted. While DNase hypersensitivity is a convenient tool for identifying relatively long linkers, micrococcal nuclease has proven to be a more versatile tool for mapping nucleosomes *in vivo* (Keene and Elgin, 1981; Noll, 1974), as extensive micrococcal nuclease digests leave the majority of nucleosomal DNA intact while extensively degrading even short linker DNAs.

In yeast, seminal studies in the Horz laboratory showed that under repressive conditions the *PHO5* promoter is associated with several nuclease-

hypersensitive sites, between which were located several “well-positioned” nucleosomes (Almer and Horz, 1986; Almer et al., 1986). A key insight from these papers was the identification of positioned nucleosomes, whose existence can be attributed to relatively little cell to cell variation in the location of the histone octamer. The contrasting case of delocalized, or “fuzzy” nucleosomes, results from nucleosomes being located at varying positions in different cells in a population and in particular from being absent in others (Small et al., 2014). After Almer and Horz’s study on yeast *PHO5*, numerous additional examples of *in vivo* nucleosome positioning were reported over the decades.

The past decade has seen modern genome-wide approaches such as tiling microarrays and deep sequencing applied to MNase-digested chromatin, yielding whole genome maps of nucleosome positioning in an ever-increasing list of organisms. Nucleosome positioning has been characterized genome-wide in major model organisms including *S. cerevisiae*, *S. pombe*, *D. melanogaster*, *C. elegans*, *A. thaliana*, and *M. musculus*, and in a number of human cell lines. In addition, many less common model and nonmodel organisms have been subject to genome-wide analysis, including 15 additional Hemiascomycete yeasts, the Japanese killifish (medaka), *P. falciparum*, *D. discoidium*, and many others (Field et al., 2009; Marx et al., 2006; Sasaki et al., 2009; Tsankov et al., 2011, 2010; Westenberger et al., 2009). The budding yeast, *S. cerevisiae*, and humans are among the best-studied species for nucleosome positioning. Not only are these species particularly well-studied, but they also exhibit dramatic differences in

multiple aspects of genomic architecture that correlate with chromatin structure, such as average gene length, number of introns, extent of higher-order chromatin structure, relative AT% vs. GC% at promoters, and many others.

Alternative methods for mapping nucleosomes *in vivo*

While MNase does not have dramatic sequence specificity, it nonetheless does have some preference on naked DNA (Dingwall et al., 1981; Hörz and Altenburger, 1981). While this sequence preference must be kept in mind when interpreting MNase-based nucleosome mapping studies, a number of partly or completely independent methods have been developed for genome-wide analysis of nucleosome positioning and/or occupancy, and in general results using these methods largely agree with MNase-based nucleosome maps.

First, because nucleosomal DNA is generally inaccessible to nuclease attack, MNase is not the only nuclease that can be used for nucleosome mapping – DNase I digestion of yeast chromatin followed by ultradeep sequencing has been successfully used to identify *in vivo* nucleosome positions (Hesselberth et al., 2009), and comparison of MNase and caspase-activated DNase maps of reconstituted chromatin templates showed nearly identical results (Allan et al., 2012). Other enzymes that act on DNA are also inhibited by nucleosomes, and several labs have made use of DNA methyltransferases to probe nucleosome positioning in isolated nuclei (Bell et al., 2010; Kelly et al., 2012) or in intact cells (Jessen et al., 2006). Here, treatment of chromatin with

methylases results in extensive methylation of linker (but not nucleosomal) DNA, which can then be interrogated by isolation of methylated DNA or by bisulfite sequencing. Nucleosomal footprints are then revealed as locations protected from cytosine methylation.

Second, a technique termed FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) relies on differential solubility of naked DNA and protein-associated DNA during phenol extraction(Lee et al., 2004; Nagy et al., 2003). Heavily protein-associated nucleosomal DNA partitions to phenol, while nucleosome-depleted DNA is preferentially recovered in the aqueous phase (Hogan et al., 2006). An alternative way to measure overall histone occupancy of DNA is to use Chromatin Immunoprecipitation (ChIP) – ChIP against histone H3 is often used to infer overall nucleosome occupancy (Bernstein et al., 2004; Fan et al., 2010), and experiments using epitope-tagged or *in vivo* biotinylated histones have also been quite successful (Mito et al., 2005).

Finally, perhaps the most exciting new technique for nucleosome mapping was reported by Widom and colleagues (Brogaard et al., 2012). Here, yeast were engineered to carry nucleosomes bearing a specific cysteine (H4S47C) located near the DNA backbone. After recovery of chromatin, a sulfhydryl-reactive copper-chelating reagent was reacted with the nucleosomal cysteine, enabling localized production of OH radicals. Subsequent mapping of DNA cleavage sites identifies genome-wide positions of nucleosomes with reported single nucleotide

resolution. This stands in contrast to nucleosome mapping that relies on MNase, as this enzyme continues to “chew away” at nucleosome ends during extended digestion protocols, leaving behind at each nucleosome position a population of nucleosome-protected DNA fragments with substantial (~10-20 bp) variability in the precise fragment ends. Overall, results obtained using ALL of these alternative methods are highly concordant with those obtained using MNase, lending strong independent support to the widespread use of MNase in nucleosome mapping studies.

In vivo nucleosome positioning patterns

As noted above, genome-wide nucleosome positioning maps have been established for at least 30 different species. Moreover, nucleosome positioning has been characterized in multiple cell types for several multicellular organisms, in a panel of genetically distinct human lymphoblastoid cell lines, and in scores of chromatin-related budding yeast mutants (van Bakel et al., 2013; Carone et al., 2014; Gaffney et al., 2012; Gkikopoulos et al., 2011; Hartley and Madhani, 2009; Mavrich et al., 2008a; Parnell et al., 2008; Schones et al., 2008; Tirosh et al., 2010; Tolkunov et al., 2011; Valouev et al., 2008; Whitehouse et al., 2007; Yen et al., 2012; Zentner et al., 2013). Across this multitude of cell types and genetic backgrounds, some surprisingly consistent themes emerge, making studies in budding yeast and human cells fairly representative.

Across all organisms studied, promoters and other regulatory elements are typically nucleosome-depleted (Radman-Livaja and Rando, 2010). In yeast, nucleosome depletion is strongest at promoters, but modest nucleosome depletion is also observed at the 3' ends of genes, even at intergenic regions located in between convergently-transcribed genes (Mavrich et al., 2008b). In humans, nucleosome depletion is also observed at other regulatory elements, such as distal enhancers (Heintzman et al., 2007; Verzi et al., 2010; He et al., 2010) or binding sites for the insulator factor CTCF (Fu et al., 2008). The region of diminished nucleosome occupancy observed at regulatory elements is variably referred to in the literature as the "nucleosome-free region" (NFR) or the "nucleosome-depleted region" (NDR), NDR being the more inclusive term.

Nucleosome depletion at promoters typically correlates with transcription rate – in all organisms studied, highly-transcribed genes are strongly nucleosome-depleted (Kelly et al., 2012; Lee et al., 2007; Schones et al., 2008; Valouev et al., 2011; Wang et al., 2012; Weiner et al., 2010). Poorly-expressed genes are fully nucleosome-occupied in human immune cells (Valouev et al., 2011), whereas they are still moderately nucleosome-depleted in budding yeast (Radman-Livaja et al., 2011) – this difference likely stems from the fact that very few genes are truly silent in yeast, and from the much more widespread role for "antinucleosomal" sequences in promoter chromatin structure in yeast relative to humans (see below).

In general, nucleosomes surrounding nucleosome-depleted regions are relatively well-positioned, but nucleosome positioning decays with increasing distance from the NDR. This observation is most clearly demonstrated at yeast promoters, where the first nucleosome downstream of a promoter (typically called the “+1” nucleosome to denote its location relative to the transcription start site) is characteristically extremely well-positioned (Figure I.2). Delocalized, or “fuzzy” nucleosomes, are primarily located near (but somewhat downstream of) the midpoints of long coding regions (Mavrich et al., 2008b; Vaillant et al., 2010; Yuan et al., 2005). Similar results hold in humans, where well-positioned nucleosomes are observed flanking NDRs at promoters, enhancers, and CTCF-binding sites. +1 nucleosome positioning is particularly strong at promoters associated with paused RNA polymerase (Schones et al., 2008), but can also be observed at a subset of promoters lacking reported polymerase pausing. As in yeast, nucleosomes located distal to regulatory elements become increasingly-poorly positioned with increasing distance, with little discernable positioning for the majority of nucleosomes located farther than ~10 positions away from a promoter, CTCF binding site, or other nucleosome-depleted region. Thus, because genes are far longer in mammals than in yeast (where the average gene is shorter than 2 kb), this means that the majority of nucleosomes are delocalized in mammalian cells, whereas the converse is true in yeast.

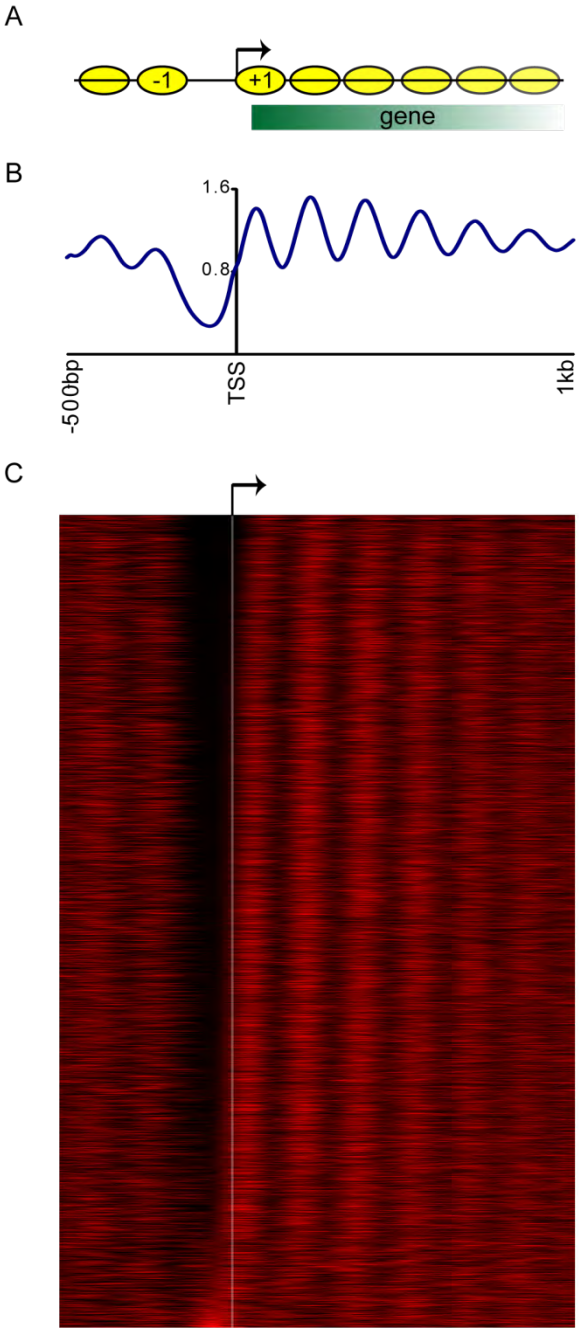


Figure I. 2: Canonical genic nucleosome architecture.

Cartoon (A), average nucleosome profile (B), and heat map (C) of nucleosome positions surrounding the transcription start site. The +1 nucleosome just downstream of the TSS is characteristically well-positioned adjacent to the nucleosome depleted promoter, which is flanked upstream by the -1 nucleosome. The conservation of nucleosome architecture across all yeast genes is evident in the heat map, where each row is a gene, aligned by the TSS.

Another major distinction between yeast and mammals is the relative abundance of introns, which are quite rare in budding yeast (~250 intron-containing genes out of ~5500 total genes) but nearly ubiquitous in mammals and many other metazoans. Curiously, nucleosome occupancy is quite different between introns and exons, with exons typically exhibiting significantly elevated nucleosome occupancy relative to introns in humans, worms, flies (Schwartz et al., 2009), fission yeast (Tilgner et al., 2009; Wilhelm et al., 2011), and others. This likely results from the generally elevated GC% found at exons, which is expected to thermodynamically favor nucleosome formation relative to the AT-rich sequences at introns (see below).

Variations on the basic theme

The averaged chromatin structure of promoters across an organism's genome is highly reproducible (Figure I.2C), with fairly similar qualitative features being observed in all organisms studied. However, there is substantial variation from gene to gene within an organism. Most notably, the extent of nucleosome depletion at promoters is highly variable, ranging from the high nucleosome occupancy observed at silenced genes in humans to an apparently complete absence of nucleosomes at highly-transcribed genes in yeast. Beyond this unsurprising link to transcription rate, genes can be broadly grouped into two classes with distinctive chromatin packaging and regulatory machinery.

In yeast, the major distinction between gene classes is that between “growth” genes, encoding the machinery involved in rapid biomass production such as ribosomal proteins, and “stress” genes, encoding cellular components such as chaperones and redox enzymes required to tolerate suboptimal conditions (Rando, 2012). This distinction between growth and stress genes is the first principal component in every genome-wide dataset we can think of. Relative to stress genes, growth genes are more often essential, exhibit more consistent protein abundance from cell to cell (Choi and Kim, 2009; Choi et al., 2008; Newman et al., 2006), are more conserved (both in copy number and sequence) over evolutionary time (Wapinski et al., 2007), are regulated by TFIID rather than SAGA (Huisinga and Pugh, 2004), are relatively insensitive to mutations in chromatin-remodeling complexes (Basehoar et al., 2004), carry more AT-rich promoters (Field et al., 2008), and have poorly conserved TATA boxes (Basehoar et al., 2004; Rhee and Pugh, 2012). In terms of chromatin structure, growth genes exhibit more strongly nucleosome-depleted promoters, whereas nucleosome depletion at stress genes is less dramatic (Field et al., 2008; Tirosh and Barkai, 2008; Weiner et al., 2010). Furthermore, growth genes are characterized by a strongly-positioned +1/-1 pair of NDR-flanking nucleosomes, whereas promoter nucleosomes at stress genes are relatively delocalized. It has been proposed that these chromatin features are at least partly responsible for the regulatory differences between the gene classes – competition between nucleosomes and TFs at stress genes likely explains why

stress genes are more reliant on chromatin regulators for proper expression, for instance (Field et al., 2008).

In mammals, the analogous distinction to the “growth/stress” axis in yeast is between “housekeeping” genes and tissue-specific genes. As in yeast, this distinction strongly correlates with sequence features – ubiquitously-expressed housekeeping genes are generally associated with CG-rich promoters (“High CpG Promoters”, or HCPs), while tissue-specific genes typically have CpG-poor (LCP) promoters (Deaton et al., 2011). As in yeast, these promoter classes differ in countless aspects of their chromatin packaging and gene regulation (Vavouri and Lehner, 2012). A curious aspect of this homology is while in yeast essential, highly-expressed genes have AT-rich promoters that are intrinsically nucleosome-depleted, in humans housekeeping genes have GC-rich promoters which are intrinsically favorable for nucleosome assembly (Tillo and Hughes, 2009; Valouev et al., 2011).

In addition to gene-to-gene variation in promoter packaging state, the internucleosomal spacing over coding regions can also vary. Indeed, the average linker length between adjacent nucleosomes differs between species (Van Holde, 1989), between cell types in multicellular organisms (Teif et al., 2012; Valouev et al., 2011), and between individual genes within a given species.

Nucleosome stability and dynamics

It has long been known that linkers differ in their susceptibility to nuclease – nucleosome-depleted regions at regulatory elements are cleaved at low levels of MNase or DNase digestion (eg are nuclease hypersensitive), whereas longer digestion times are required to cleave linkers located within gene coding regions. Similarly, nucleosomes differ in their stability during long MNase digestions (Bryant et al., 2008). Recent genome-wide studies have characterized this behavior in some detail via deep sequencing of mononucleosomal DNA isolated after varying length MNase digestions (Weiner et al., 2010; Xi et al., 2011). In general, the length of DNA remaining after MNase digestion decreases with increasing digestion time or MNase concentration, as MNase slowly chews away the ends of nucleosomal DNA. In addition, a subset of “fragile” nucleosomes, typically occurring at nucleosome-depleted regions such as promoters and 3’ ends of genes, are observed only under relatively mild digestion conditions, when the majority of chromatin remains in partially-digested fragments consisting of two or more nucleosomes. Fragile nucleosomes are highly correlated with localization data for the H2A.Z histone variant, with a large number of 5’-biased histone modifications including many histone acetylation marks, and with sites of rapid replication-independent nucleosome turnover (Dion et al., 2007; Rufiange et al., 2007).

How are nucleosome positions established *in vivo*?

The surprising order observed in genome-wide nucleosome positioning studies – particularly in organisms with compact genomes – has further motivated an already sizeable community to attempt to understand the mechanistic basis underlying the establishment of nucleosome positions. It has long been known that nucleosome formation is thermodynamically preferred over some sequences relative to others (Drew and Travers, 1985), raising the possibility that chromatin architecture is "programmed" by the genome's sequence. In contrast, it has also been clear for decades that proteins from transcription factors to RNA polymerase to the broad group of ATP-dependent chromatin remodelers can influence nucleosome positioning. While it is abundantly clear that both *cis* (DNA sequence) and *trans* (remodelers, etc.) factors have some influence over chromatin architecture, the past 7-8 years have seen a lively debate regarding the relative contributions of these factors to *in vivo* nucleosome positioning.

Sequence Biases in Nucleosome Positioning – favorable sequences

As expected for a whole-genome packaging factor, histones do not have sequence-specific DNA-binding domains. Most interactions between histone proteins and DNA observed in the nucleosome crystal structure are between residues that penetrate into the minor groove, and the phosphodiester backbone (Luger et al., 1997). Despite this, early studies showed that the sea urchin 5S rRNA sequence, which is associated with a well-positioned nucleosome *in vivo*,

was sufficient to direct a positioned nucleosome in *in vitro* reconstitutions where no other proteins are present to influence nucleosome assembly (Simpson and Stafford, 1983), and mutations in the sequence were shown to interfere with appropriate positioning (FitzGerald and Simpson, 1985). These and many related studies identified a role for sequence in modulating the affinity of a stretch of DNA for histone proteins and thus in the preferred placement of nucleosomes.

Seminal studies from Drew and Travers investigated the role of DNA flexibility in constraining nucleosome positioning – in order to be incorporated into a nucleosome, DNA must dramatically bend to wrap around the histone octamer, and differences in DNA flexibility thus can alter the affinity of a given DNA sequence for the histone octamer. Drew and Travers observed a rotational preference for DNA formed into a nucleosome *in vitro*: short AT runs positioned with their minor grooves facing inwards, with GC runs positioned facing outwards (Drew and Travers, 1985). Cloning nucleosomal DNA from chicken erythrocytes further revealed a ~10.15 bp periodicity of di- and tri- nucleotides in nucleosomal DNA, with AT dinucleotides prevalent where the minor groove faces inward towards the nucleosome core, and GC occurring where the DNA minor groove faces away from the nucleosome (Satchwell et al., 1986). The preferential positioning of AT-rich stretches with minor grooves facing inwards, confirmed in a plethora of subsequent mapping studies (Brogaard et al., 2012), has been ascribed to the fact that AT-rich minor grooves are narrower than those of GC-rich sequences, enabling them to better suffer the compression of the minor

groove (Figure I.1) required by the smaller internal radius (Chua et al., 2012; Morozov et al., 2009; Olson and Zhurkin, 2011; Tolstorukov et al., 2007). Indeed, even in the absence of histones, the Drew and Travers circular DNA template formed with A+T stretches in the internally facing minor groove (Drew and Travers, 1985). Beyond this indirect readout of shape, recent work suggests that AT dinucleotide periodicity might additionally reflect improved thermodynamics of electrostatic interactions between histone arginine side chains and the narrowed minor groove (West et al., 2010).

Armed with the knowledge that sequence could direct nucleosome positioning, Widom and colleagues undertook an extensive series of selection studies to isolate exceptionally strong nucleosome positioning sequences, resulting in the now-famous “Widom601” nucleosome positioning sequence (Lowary and Widom, 1998). This sequence has been used in thousands of subsequent studies requiring nucleosomes positioned at precisely-defined locations on a template *in vitro*, and therefore provides the foundation for much of our understanding of nucleosome biochemistry (although the unusually strong affinity of this sequence for histones is worth keeping in mind when extending interpretations of such studies to more typical nucleosomes encountered in the genome). The Widom601 and many other sequences selected for preferential nucleosome incorporation generally show evidence of being inherently bent (or having a greater ability to bend) – these sequences tend to exhibit strong periodic spacing of AT dinucleotides (Lowary and Widom, 1998), as is seen in

natural nucleosome sequences (Brogaard et al., 2012; Satchwell et al., 1986). When the 10 bp phasing of AT dinucleotides is disrupted or they are removed, the relative affinity of these sequences for nucleosomes is reduced (Segal et al., 2006), suggesting that this characteristic is necessary for strongly positioning nucleosomes. Several groups have taken advantage of the periodicity of dinucleotides seen in nucleosomal DNA to develop models that are able to predict nucleosome positions and occupancy to some extent based on sequence (Ioshikhes et al., 1996, 2006; Segal et al., 2006).

However, while early computational models based on AT-periodicity claimed to provide a modest improvement over random models, these models dramatically fail to predict major features of *in vivo* nucleosome positions, such as promoter nucleosome depletion. In subsequent years, *in vitro* reconstitution of nucleosomes found no strong signal for nucleosome positioning at the +1 nucleosome, rather revealing the major signal to be nucleosome depletion over promoters in the yeast genome (Kaplan et al., 2009; Zhang et al., 2009) (Figure I.3A). This observation is consistent with the improvement of computational models by the incorporation of nucleosome depleted sequences (Field et al., 2008). Thus, the major sequence contribution to chromatin architecture in yeast, at least at ~10 bp resolution, is the role of antinucleosomal sequences at promoters.

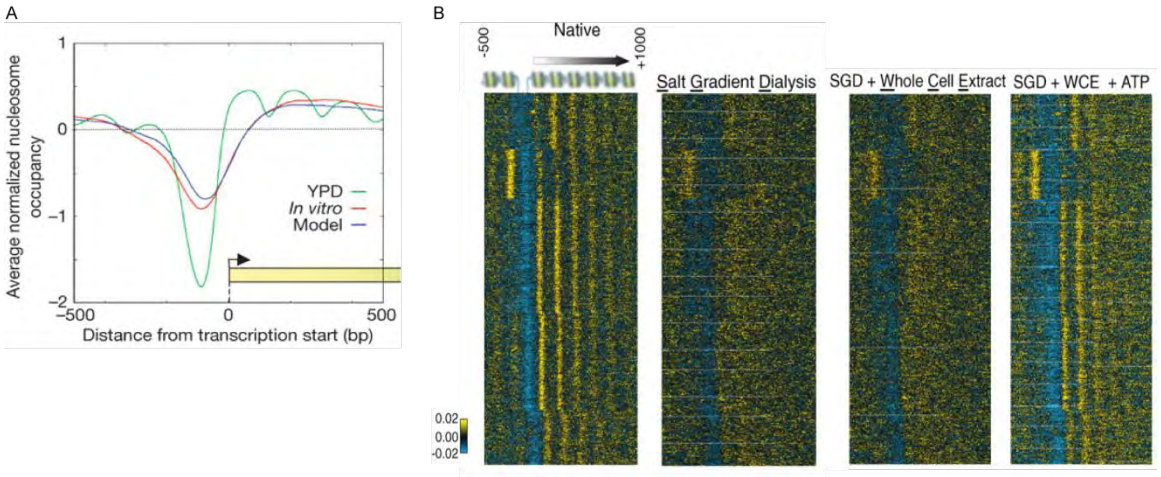


Figure I. 3: Nucleosome positioning by sequence and trans-acting factors.

A) *In vitro* reconstituted nucleosome positions and sequence-based models of nucleosome positions support nucleosome depletion in *cis*.
B) *in vitro* reconstitution with yeast whole cell extract shows importance of ATP-dependent chromatin remodellers in establishing native chromatin structure.

Panel A reprinted by permission from Macmillan Publishers Ltd: Nature, Kaplan et al., 2008.

Panel B from Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Körber, P., Pugh, B.F., (2011) A Packing Mechanism for Nucleosome Organization Reconstituted Across a Eukaryotic Genome, Science, 332: 6032. Reprinted with permission from AAAS.

Sequence Biases in Nucleosome Positioning – antinucleosomal sequences

In addition to the role of AT-periodicity in DNA bending, Travers and colleagues also noted that homopolymeric stretches such as PolyA, which are intrinsically stiff, might be preferentially excluded from nucleosomes and would likely be overrepresented in linker DNA. Indeed, the major signal in genome-wide nucleosome reconstitution studies comes from nucleosome depletion over sequences rich in poly(dA:dT) runs (Figure I.3A). Early studies suggested that

long poly(dA:dT) tracts would not assemble into nucleosomes (Rhodes, 1979; Simpson and Künzler, 1979), although later studies were able to assemble these tracts into nucleosomes, albeit with reduced affinity relative to random sequences (Anderson and Widom, 2001; Puhl and Behe, 1995; Puhl et al., 1991; Satchwell et al., 1986). These *in vitro* studies were complemented by classic *in vivo* experiments in yeast from Iyer and Struhl which showed that nucleosome depletion at the *HIS3* promoter was strongly correlated with the length of the poly(dA:dT) tract (Iyer and Struhl, 1995). Subsequent work showed that the extensive nucleosome depletion associated with promoters in yeast can be recapitulated via *in vitro* reconstitution of yeast genomic DNA (Sekinger et al., 2005), and early *in vivo* mapping studies in yeast showed a very strong correspondence between PolyA elements and promoter nucleosome depletion (Yuan et al., 2005). Genome-wide reconstitution studies confirmed that the major signal for DNA sequence in establishing nucleosome occupancy was that of PolyA elements (Kaplan et al., 2009; Zhang et al., 2009) (Figure I.3A). Computational analyses of *in vivo* nucleosome mapping data further support a dominant role for antinucleosomal PolyA runs as key sequence determinants of the chromatin landscape (Field et al., 2008; Yuan and Liu, 2008), and in fact a model based solely on GC% was nearly as successful as far more complex computational models at predicting nucleosome occupancy from genomic sequence (Tillo and Hughes, 2009).

The major role of promoter PolyA elements as a sequence determinant of nucleosome organization in yeast is now well-established. Importantly, nucleosome depletion over PolyA elements is a quantitative feature, and scales with the number and length of PolyA elements at a given promoter. Because homopolymeric runs such as PolyA runs are labile – they suffer relatively rapid sequence expansions and contractions over generations of growth – PolyA length elements provide a mechanism for evolutionarily rapid control of a gene's promoter accessibility, expression level, and regulatory program. PolyA strength at many promoters is polymorphic between different budding yeast strains, and changes in PolyA length are observed at target promoters even during short laboratory evolution experiments (Vinces et al., 2009). Over longer time scales, PolyA gain and loss contributes to the evolution of gene regulatory programs. Most dramatically, in Hemiascomycota fungi that preferentially generate energy via respiration (those that diverged prior to a whole genome duplication event), promoters of genes involved in respiration (mitochondrial proteins, TCA cycle genes, etc.) are associated with long PolyA tracts and are strongly nucleosome-depleted *in vivo* and *in vitro* (Field et al., 2009; Tsankov et al., 2010). In species that preferentially obtain energy via fermentation, these genes have lost PolyAs and gain nucleosome occupancy, and exhibit other characteristics of "stress", rather than "growth", genes.

Moreover, the extent to which an organism utilizes antinucleosomal sequences such as PolyA tracts varies across eukaryotic species. Among

Hemiascomycota, for example, PolyA was shown to be used to a similar extent (PolyA tract length, number of genes associated with strong PolyA tracts) in a number of species, but PolyA tracts were found to be relatively short and infrequent in the halophilic yeast *D. hansenii* (Tsankov et al., 2011, 2010). More distantly, a number of larger eukaryotes appear to use PolyA tracts sparingly, as for example in mammals housekeeping genes are instead associated with the GC-rich CpG island promoters. These promoter sequences strongly favor nucleosome assembly (Tillo et al., 2010; Valouev et al., 2011), in contrast to the nucleosome depletion programmed at yeast growth genes. Interestingly however, there is some evidence in metazoans that GC-rich promoters are "paradoxically" accessible – for example, upon p53 induction, those p53 motifs that are located in GC-rich promoters and are nucleosome-occupied become p53-bound in preference to equally-strong motifs found in low GC-promoters (Lidor Nili et al., 2010). In this and other cases, it is plausible that "overpackaging" with nucleosomes may prevent compaction into higher-order chromatin structures such as 30 nm fiber, enabling TF access to genomic loci despite high nucleosome occupancy. No doubt future studies will provide insights into the forces shaping the evolution of overall genomic nucleotide composition as well as the role for local deviations from the genomic average in dictating promoter packaging and access.

Taken together, computational analyses of *in vivo* nucleosome mapping data and experimental study of *in vitro* nucleosome reconstitutions reveal two

mechanisms by which DNA sequence affects the chromatin landscape. A major force in relative nucleosome occupancy is AT% versus GC%, with high AT% at yeast promoters associated with nucleosome depletion, and the high GC% at mammalian promoters driving high intrinsic nucleosome occupancy. In addition, a multitude of studies indicate intrinsically high affinities of sequences with 10 bp AT-dinucleotide periodicity for the histone octamer. While these sequences do not direct strong nucleosome positioning in *in vitro* reconstitutions, they more likely direct local rotational positioning; that is, the orientation of the minor groove in relation to the nucleosome core. In other words, we may consider AT dinucleotide periodicity as driving a “toothed” local thermodynamic landscape, where extrinsic factors drive positioning to +/- 5 bp, with the local landscape directing the precise sequences that bend inwards vs. outwards and thereby influencing major groove sequence access. Indeed, single bp precision nucleosome mapping suggests that this rotational positioning occurs throughout yeast coding regions (Brogaard et al., 2012). It will be interesting in the future to see the functional consequences of such rotational positioning.

Packing effects in nucleosome positioning – “statistical positioning” and related models

While sequence cues clearly affect nucleosome occupancy, it is less clear why the majority of nucleosomes are well-positioned in organisms with compact genomes. In general, nucleosomes surrounding regulatory elements are well-

positioned, with nucleosome “fuzziness” increasing with increasing distance from such elements. Broadly speaking, these observations are consistent with Kornberg and Stryer’s “statistical positioning” hypothesis (Kornberg and Stryer, 1988). This hypothesis holds that even if there is no intrinsic preference for nucleosomes to form at a given position along a stretch of DNA, if nucleosomes are fixed at the ends of a given stretch (by transcription factors, sequence cues, etc.) then intervening nucleosomes could exhibit strong positioning as a result of straightforward packaging considerations. An analogy we like to use (Rando and Chang, 2009) is that of a can of tennis balls – a single tennis ball in a can could occupy many positions throughout the can, but three tennis balls in a can will all be well-positioned. Many observations are consistent with this – in human cell lines, phased arrays of nucleosomes are often found surrounding nucleosome-depleted loci, but are lost in cell types or in genetic backgrounds with diminished nucleosome depletion at the putative boundary (Gaffney et al., 2012; Wang et al., 2012). Moreover, when yeast are subject to global nucleosome depletion (Gossett and Lieb, 2012), nucleosome positions appear to be less sharply defined, consistent with relaxed packing constrains. However, nucleosome occupancy rather than distribution appears to be affected particularly at sequences that are less predisposed to wrap nucleosomes (Celona et al., 2011; Gossett and Lieb, 2012), which indicates that chromatin packing is not exclusively mediated by statistical positioning.

In a detailed analysis of this phenomenon, Vaillant et al sorted yeast genes by the distance from the first (+1) to the last (+N) nucleosome associated with a coding region (Vaillant et al., 2010), finding that the peak of nucleosome fuzziness typically occurred roughly 2/3 of the way along the length of the gene. Interestingly, the authors also noted that for gene lengths that were an integral repeat of the internucleosomal repeat length exhibited “crystalline” nucleosome arrays with well-positioned nucleosomes, while nonintegral repeats were associated with fuzzier nucleosome arrays. In other words, tennis balls in a 3 or 4 ball-length can are all well positioning, but the balls in a 3.5 ball length can have extra room to bounce around. Computational models also support statistical positioning, as average nucleosome positioning over yeast genes can be explained using a “Tonk’s gas” one-dimensional gas formalism(Möbius and Gerland, 2010; Möbius et al., 2013). Interestingly, in these models the decay of nucleosome positioning is best-explained with nucleosomes downstream of a promoter packing against a positioned +1 nucleosome, but with upstream nucleosomes being constrained instead by nucleosome-disfavoring sequences at promoters. This asymmetry may help explain why the greatest nucleosome fuzziness over coding regions occurs downstream of genic midpoints.

Together, these observations lend extensive support to the idea that statistical positioning accounts for the surprising order observed for nucleosomes in compact genomes, although it is worth noting that some observations do not support this model (see, eg, (Zhang et al., 2011)).

Nucleosome positioning: *trans*-acting factors

Altogether, there is little question that sequence features play some role in influencing nucleosome *occupancy in vivo*, it is quite clear that precise nucleosome *positioning in vivo* results from *trans*-acting factors, primarily proteins. This is apparent from the results of reconstitution assays in which yeast genomic DNA is incorporated into nucleosomes *in vitro* – as noted above, DNA sequence alone is sufficient to direct nucleosome depletion at many promoters (Kaplan et al., 2009; Korber and Hörz, 2004; Sekinger et al., 2005; Zhang et al., 2009), but beyond promoter nucleosome depletion very few features of nucleosome positioning are recovered in these assays. For instance, DNA sequence is quite insufficient to accurately direct positioning of the well-positioned +1 nucleosome in such studies. Moreover, several groups have found that even the strongest nucleosome positioning sequence, the “Widom601” sequence, fails to direct nucleosome positioning when engineered into intact cells (Gracey et al., 2010; Perales et al., 2011). The disconnect between sequence preferences and *in vivo* positioning is particularly pronounced in humans, where the generally GC-rich promoters are sites of preferential nucleosome incorporation *in vitro* (Tillo and Hughes, 2009; Valouev et al., 2011), yet are in fact nucleosome-depleted *in vivo*.

In yeast, two approaches provide particularly strong evidence for a general role for *trans*-acting factors in nucleosome positioning. First, Korber and

colleagues have developed an elegant extract-based reconstitution to identify factors required for appropriate *in vivo* nucleosome positioning. Originally, they used this system to show that while *PHO5* chromatin could not be recapitulated by salt dialysis to deposit histones on DNA *in vitro*, addition of yeast whole cell extract was sufficient to correctly assemble this gene's chromatin architecture *in vitro* (Korber and Hörz, 2004). Subsequent genome-wide analysis showed that yeast extract could broadly help assemble nucleosomes on the yeast genome in patterns closely resembling their true *in vivo* positions (Zhang et al., 2011) (Figure I.3B). This biochemical approach promises to enable powerful future studies such as fractionation of activities required for proper nucleosome positioning over specific promoters (Wippo et al., 2011).

A second class of approach to unraveling *cis* and *trans* determinants of chromatin structure is primarily genetic. For example, the Kruglyak lab pioneered the approach of crossing two yeast strains (BY and RM), and analyzing hundreds of haploid segregants to dissect *cis* and *trans* genetic effects on transcription (Brem et al., 2002). This approach was recently extended to an analysis of open chromatin (analyzed using FAIRE) in 96 haploid segregants, with a key insight being that the vast majority (~90%) of differences between open chromatin between these strains was linked to *trans*-acting loci such as chromatin regulators (Lee et al., 2013). A related genetic approach to generally separate *cis* and *trans* effects on nucleosome positioning is to analyze chromatin structure over a given sequence carried in alternative species' genomes. For

instance, Struhl and colleagues showed that for the budding yeast *HIS3* promoter, nucleosome depletion was maintained over this promoter when it was inserted into the *S. pombe* genome, but only 2/7 nucleosomes were correctly positioned (Sekinger et al., 2005). Using hybrids between the closely-related *S. cerevisiae* and *S. paradoxus*, Tirosh and Barkai showed that differences in poly(dA:dT) abundance could explain in most cases those chromatin differences that were linked in *cis* to genomic sequence (Tirosh et al., 2010).

Taken together, results from *in vitro* reconstitution experiments with and without cell extracts, and from experiments using heterologous sequences in different cells, demonstrate that the majority of nucleosomes depend on *trans*-acting factors for correct *in vivo* positioning. The three major classes of *trans*-acting factor that can affect nucleosome positioning are transcription factors, RNA polymerase, and ATP-dependent remodelers.

ATP-dependent chromatin remodeling enzymes

As noted above, yeast genomic DNA can be assembled *in vitro* into nucleosomes that reasonably match *in vivo* positions *only* when using whole cell extract. This assembly requires ATP hydrolysis, pointing to the general importance of a major class of chromatin regulatory factors – the ATP-dependent chromatin remodelers (Clapier and Cairns, 2009). ATP-dependent chromatin remodelers are defined by the presence of an ATPase subunit homologous to the yeast Swi2/Snf2 (Peterson and Tamkun, 1995; Winston and Carlson, 1992)

protein – in budding yeast, 17 Snf2 homologs exist, while humans carry 53. Snf2 homologs typically occur in multisubunit complexes, and these complexes exhibit a wide variety of activities on chromatin substrates. They share in common the use of ATP hydrolysis to disrupt histone-DNA interactions and thereby “loosen” DNA on the surface of the octamer. Depending on the ATPase in question and its associated subunits, eventual outcomes of the remodeling reaction can include nucleosome “sliding” (eg lateral movement to a new position without histone loss), partial or complete nucleosome eviction, and alterations in the octamer composition such as histone dimer exchange (Figure I.4). It is worth noting that these outcomes are extensively entangled *in vivo* (Tomar et al., 2009). For example, the Ino80 complex appears to have a major role in H2A/H2A.Z exchange, yet its deletion exhibits altered nucleosome positioning *in vivo* (van Bakel et al., 2013; Tirosh et al., 2010; Yen et al., 2012). However, whether nucleosome locations in this mutant are altered directly by Ino80 sliding activity, which has been observed *in vitro* (Udugama et al., 2011), or results from H2A.Z’s ability to interfere with or enhance other ATP-dependent remodelers (Li et al., 2005) remains unknown.

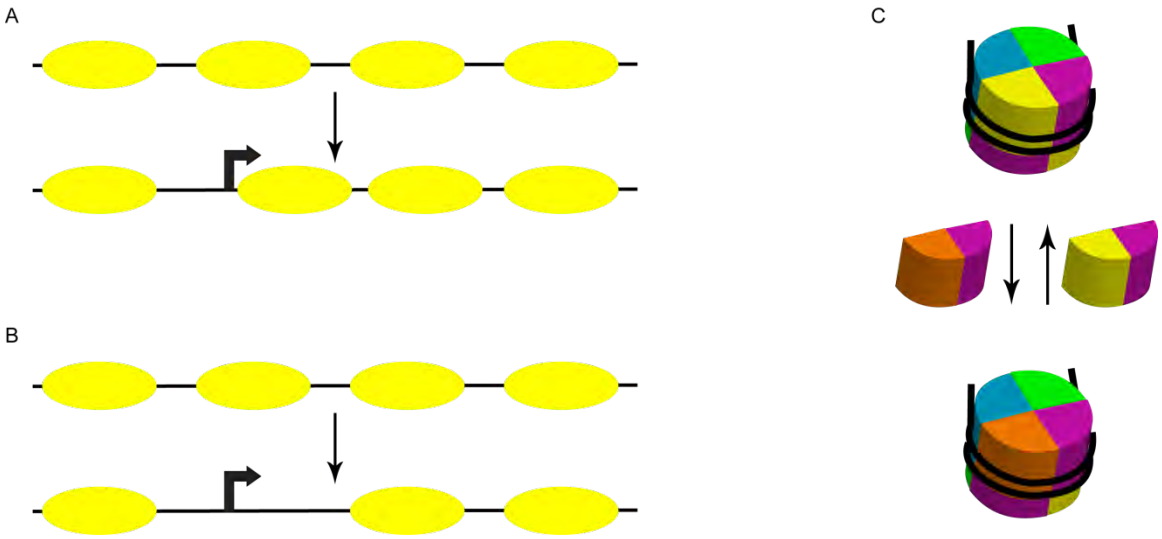


Figure I. 4: ATP-dependent chromatin remodellers affect chromatin structure in several ways.

A) Nucleosome sliding by remodellers, such as Isw and Chd class remodellers, can expose or occlude DNA binding sites.
B) SWI/SNF class remodellers are capable of evicting nucleosomes to expose underlying DNA.
C) Swr1/Ino80-class remodellers exchange H2A/H2B and H2A/Z/H2B dimers to alter chromatin composition.

In general, Swr1/Ino80-class remodelers primarily act to exchange H2A/H2B and H2A.Z/H2B dimers with one another (Figure I.4C) (although Ino80 can also slide nucleosomes *in vitro*), Isw-class remodelers play roles in lateral nucleosome sliding (Figure I.4A), and Snf2-class remodelers alter nucleosome occupancy by eviction/reassembly of histone octamers (Figure I.4B). *In vitro*, different remodeling complexes can have quite distinct activities on the same

templates; for instance Isw1a, Isw2, and Chd1 tend to centrally position nucleosomes on short DNA fragments, while Isw1b and human SWI/SNF are capable of moving nucleosomes toward the end points of DNA (Bouazoune et al., 2009; Stockdale et al., 2006). As these remodelers serve to generally loosen DNA-histone contacts, this may allow the nucleosome to sample relatively unfavorable sequence locations, with accessory subunits in the various complexes or specific domains in the ATPases themselves serving to stabilize different remodeling outcomes (end vs. central, for example). For instance, Isw1, Isw2, and Chd1 exhibit particularly strong binding to nucleosomes carrying extranucleosomal DNA, which may be involved in their tendency to centre nucleosomes (Stockdale et al., 2006). Indeed, alterations in the DNA binding abilities of Chd1 have been able to motivate directed movement of a nucleosome toward the DNA binding site (McKnight et al., 2011). Remodeling complexes not only bind to nucleosomes and associated DNA, but can also carry domains that are regulated by covalent histone modifications or specific histone variants. Thus, combinations of DNA and histone binding domains may serve to target the remodeler or may modulate its activity at a particular site, serving to determine the directionality characteristics of nucleosome movement (Clapier and Cairns, 2009).

In vivo, chromatin remodeling complexes often appear to subvert the genomic landscape of favorable/unfavorable nucleosome positions. As a key example, now-classic work from Whitehouse and Tsukiyama showed that the

Isw2 repressor functions to position nucleosomes over intrinsically unfavorable sequences (Whitehouse and Tsukiyama, 2006; Whitehouse et al., 2007). More generally, in a study of four ATP-dependent remodelers – ISWI, (P)BAP, INO80, and NURD – in Drosophila S2 cells found that all four of these remodelers (which bound to different genomic locations) acted to subvert intrinsic sequence preferences for nucleosome occupancy (Moshkin et al., 2012). Here, ISWI was localized over predicted nucleosome-favoring sequences where it acted to reduce nucleosome occupancy, whereas the other three remodelers all bound to nucleosome-disfavoring sequences and stabilized nucleosomes. Similar results are found in many additional yeast mutants, where nucleosome positions in several remodeler deletions better match intrinsic sequence preferences than do wild-type nucleosome positions (van Bakel et al., 2013). It is worth noting that nucleosome remodelers do not universally counteract sequence preferences, as for example it has been reported in yeast that SWI/SNF functions to clear nucleosomes from intrinsically unfavorable sequences and thereby *reinforces* sequence preferences (Tolkunov et al., 2011).

In yeast, a number of studies have examined the effects of deletion of most ATP-dependent remodelers, with data currently available for mutations in the RSC complex, Isw2, SWI/SNF, Isw1a and Isw1b complexes, Ino80 and Swr1, and Chd1 (van Bakel et al., 2013; Gkikopoulos et al., 2011; Hartley and Madhani, 2009; Parnell et al., 2008; Tirosh et al., 2010; Tolkunov et al., 2011; Whitehouse et al., 2007; Yen et al., 2012), as well as several double and triple

mutants (Gkikopoulos et al., 2011). The most dramatic phenotype observed came from this last study, in which triple deletions lacking Isw1, Isw2, and Chd1exhibited reasonable positioning of the +1 and +2 nucleosomes over coding regions, but nearly complete disorganization of all downstream nucleosomes (Gkikopoulos et al., 2011). A similar effect is seen in *S.pombe* deletion mutants of its two Chd1 orthologs (Pointner et al., 2012). The generation of a normal nucleosome ladder upon gel electrophoresis would suggest that internucleosome spacing is actually maintained within a gene/cell, but that the anchor point of the nucleosome array varies from cell to cell. Further extending this study to single molecule mapping using cytosine methylation (Jessen et al., 2006) would be extremely informative.

In general, mutant and localization studies carried out under favorable growth conditions find that specific remodelers tend to exhibit consistent directional effects. For example, Isw2 and the Isw1a complex primarily shift nucleosomes at the 5' ends of genes from 3' to 5', towards the NDR – in other words, loss of either of these complexes results in +1 and +2 nucleosomes shifting downstream into the coding regions of target genes (Figure I.5). In contrast to Isw2 and Isw1a, Isw1b, Chd1, SWI/SNF, and RSC, all on average push 5' nucleosomes downstream (Figure I.5). The opposite behavior is generally observed on the other side of the NDR – Isw2, for example, shifts both NDR-flanking nucleosomes towards the NDR. A striking observation from these localization and deletion studies is that ATP-dependent remodeling enzymes

often “act at a distance.” In other words, remodeling enzymes such as Isw2 or Chd1 are primarily localized at NDR-flanking nucleosomes (Yen et al., 2012; Zentner et al., 2013), yet their deletions can affect nucleosome positions several positions away, within gene bodies. This may reflect packing effects over coding regions, in which a given remodeler directly acts on, say, the +1 nucleosome, with effects on +2 and +3 nucleosomes occurring downstream as these nucleosomes are packed against a shifted boundary.

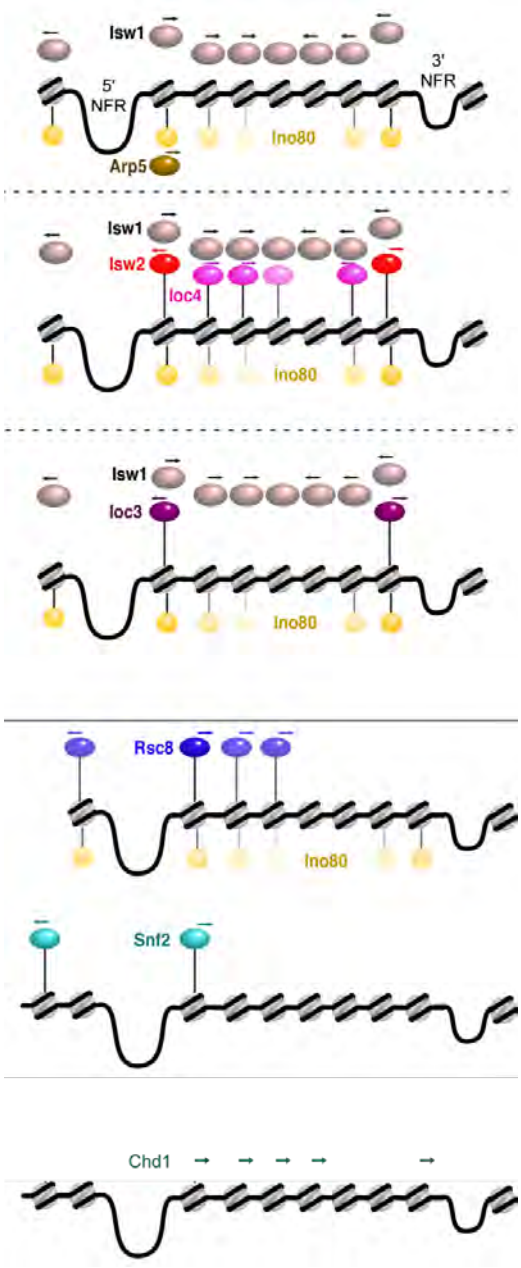


Figure I. 5: Chromatin remodeler activity is targeted to different areas of the gene.

Cartoons show the localization of chromatin remodeler with colored elipses above nucleosomes, while arrows show the directionality of chromatin remodeling, based on alterations of nucleosome positions in deletion mutants.

Adapted from Cell 149, Yen, K., Vinayachandran, V., Bata, K., Körber, R.T., & Pugh, B.F., Genome-wide Nucleosome Specificity and Directionality of Chromatin Remodelers, 1461-1473, 2012 with permission from Elsevier

General regulatory factors and other transcription factors

Even for promoter nucleosome depletion, *in vitro* reconstitutions do not perfectly recapitulate *in vivo* data– on average, yeast promoters are far more nucleosome-depleted *in vivo* than *in vitro*. This is of course far more pronounced for nucleosome depletion at promoters in other species such as humans, where extrinsic factors must reverse the nucleosome occupancy programmed by the relatively high GC%. The specific nucleosome depletion seen *in vivo* likely stems from the large number and variety of proteins that are typically associated with active promoters.

It has long been understood that transcription factors can compete with histones for binding to the same DNA sequence. For instance, Workman and Kingston showed two decades ago that binding of the Gal4p transcription factor can evict nucleosomes from a promoter (Workman and Kingston, 1992; Workman et al., 1991) by making the histones more available to acceptors such as histone chaperones (Owen-Hughes and Workman, 1996). But the relationship between transcription factors and nucleosomes is somewhat complex – nucleosomes generally inhibit binding of transcription factors to their motifs (see below), and for many transcription factors deletion studies show minimal effects on nucleosome occupancy of their binding sites. A number of features underlie the complexity of this relationship. First, it appears that transcription factors differ in their ability to invade, or displace, a nucleosome *in vivo* (Yu and Morse, 1999)

– these differences may result from differences in protein abundance or the ability to recruit chromatin remodelers, among other things. Second, as noted above, transcription factors are more able to bind to nucleosomal DNA located near the entry/exit points, where thermal breathing is more pronounced – the time scale for DNA exposure varies from ~0.25 second unwrapping times near to DNA entry-exit points to ~10 minute unwrapping times for DNA at the dyad axis (Tims et al., 2011). Third, *in vivo* promoter architecture no doubt alters the impact of transcription factor deletion – a TF binding site located next to a long PolyA sequence is more likely to remain nucleosome-depleted in the absence of the TF than is an isolated TF binding site.

In yeast, a specific family of transcription factors such as Abf1, Reb1, and Rap1, known as “General Regulatory Factors (GRFs)” (Buchman and Kornberg, 1990), has particularly robust activity in nucleosome eviction. Interestingly, while GRFs have some transactivation potential, more often they appear to facilitate gene activation by enabling the binding of other TFs with strong transactivation domains to a given promoter (Buchman and Kornberg, 1990; Yarragudi et al., 2004; Yu and Morse, 1999). The antinucleosomal capabilities of GRFs manifest in three distinct assays. First, shifting conditional GRF mutants (which are usually essential) to the restrictive temperature results in increased nucleosome occupancy over the GRF’s binding sites (Badis et al., 2008; van Bakel et al., 2013; Ganapathi et al., 2011; Tsankov et al., 2011). Second, incorporating a GRF binding motif into a heterologous location (such as the middle of a coding

region for a nonessential gene) can direct nucleosome depletion, although this appears to require at least a short PolyA stretch in addition to the GRF motif and to be mediated by RSC activity (Hartley and Madhani, 2009; Raisner et al., 2005). Finally, GRFs can be identified bioinformatically by comparing *in vivo* nucleosome occupancy to nucleosome occupancy obtained from *in vitro* reconstitutions. For example, Kaplan et al compared *in vitro* and *in vivo* nucleosome occupancy over all possible 7mer sequences, finding the expected strong nucleosome depletion over AAAAAAA (Kaplan et al., 2009). However, a number of sequences exhibit divergent nucleosome occupancy, with dramatic nucleosome depletion observed *in vivo* despite no intrinsic antinucleosomal activity *in vitro*. These sequences correspond to binding sites for the GRFs. This same approach has also been successfully extended to the genomes of other hemiascomycetes in order to predict and later validate novel GRF binding sites in other species (Tsankov et al., 2011, 2010).

In another interesting connection between the transcription machinery and nucleosome positioning, it is typically found that the first nucleosome downstream of a promoter, the +1 nucleosome, is generally relatively well-positioned, with a stereotyped positioning with respect to genic transcription start sites (TSSs) and with respect to basal transcription factors such as TFIID (Rhee and Pugh, 2012). The mechanistic basis for the stereotyped positioning of the +1 nucleosome in many organisms remains imperfectly understood. The contention that this occurs via intrinsic sequence effects (Segal et al., 2006) is clearly

incorrect, as *in vitro* nucleosome reconstitutions using only salt dialysis reveal no evidence whatsoever for +1 nucleosome positioning (Kaplan et al., 2009; Zhang et al., 2009, 2011), and even in the presence of yeast extract and ATP the positioning of the +1 nucleosome does not match *in vivo* positions.

At present, a number of pieces of evidence point toward a role for some component of the basal transcription machinery in positioning the +1 nucleosome. In CD4+T cells, binding of Pol II, regardless of active transcription, appears to play a role in nucleosome positioning, but the position of the +1 nucleosome appears to be further influenced by active elongation of Pol II in these cells (Schones et al., 2008). Similar results are obtained in flies, where Pol2 pausing has dramatic effects on nucleosome positioning – genes with high levels of paused Pol2 exhibit low levels of downstream nucleosome occupancy, and upon release of pausing by NELF knockdown nucleosome occupancy is increased over the transcription start site, specifically at genes that experience a high pause index (Gilchrist et al., 2010). This suggests that these highly paused genes are regulated by competition between the nucleosome and transcriptional machinery as well as by negative elongation regulators. Clearly nucleosome positioning and transcription are intimately linked, however the order of events in the establishment of the well-positioned +1 nucleosome and the transcription start site are not yet elucidated.

Effects of RNA polymerase on chromatin structure

Outside of DNA polymerase, the *trans*-acting protein that acts at the greatest fraction of the genome is likely RNA polymerase, as it is known in many organisms that the majority of the genome is transcribed – in yeast, 2/3 of the genome codes for proteins, while in humans a huge amount of noncoding transcription has been uncovered in the last decade. Moreover, as noted above, transcription at protein-coding genes is correlated with chromatin structure – highly-transcribed genes typically exhibit greater promoter nucleosome depletion, and shorter average internucleosomal spacing downstream, than poorly-transcribed genes.

Two general approaches reveal the effects of RNA polymerase on nucleosome positioning – either changing transcription rates *in vivo* by changing growth conditions, or more directly by analyzing chromatin structure in the absence of transcription, either *in vivo* or *in vitro*. A large number of studies have examined the effects of altering transcriptional programs on nucleosome positioning, starting with Horz's classic work on the *PHO5* promoter. In the genomics era, changes in nucleosome positioning/occupancy in yeast has been studied in response to carbon source shifts, various stress conditions, and during the meiotic program, while studies in larger eukaryotes include heat stress in flies and TCR activation in human CD4+ T cells. Given that promoter nucleosome occupancy *in vivo* typically anticorrelates with transcription rate, it is therefore

unsurprising that this relationship is dynamic – many genes repressed during a given transcriptional response exhibit nucleosome gains when repressed. That said, it is worth noting that a large number of genes can be found in any gene induction/repression response that exhibit strong mRNA abundance changes without any apparent change in promoter nucleosome positioning/occupancy.

Of course, alterations in chromatin observed during an *in vivo* transcriptional response might reflect the direct action of RNA polymerase, but could just as easily result from changes in TF binding or chromatin remodeling activities. Slightly more direct than inducing or repressing gene transcription are studies on chromatin state in the presence or complete absence of transcription. In most multicellular organisms a number of cell divisions occur in the early embryo prior to activation of the zygotic genome, and in flies nucleosomes have been mapped before and after this stage (the mid-blastula transition) (Moshkin et al., 2012). Unsurprisingly, in both of these cases nucleosomes in the untranscribed state exhibited better correlations with the pure thermodynamic preferences of genomic sequences for nucleosome incorporation, while transcription clearly is capable of subverting these intrinsic preferences.

In vitro transcription assays on a chromatin template have shown that RNA polymerase passage results in upstream trafficking of histone proteins. Here, multiple groups have found that some RNA polymerases are capable of shifting histones in a direction retrograde to polymerase passage, apparently as

a result of polymerases pushing a bubble of DNA around the octamer (Bintu et al., 2011; Kulaeva et al., 2009; Studitsky et al., 1994, 1997). The ability of Pol2 in particular to shift nucleosomes upstream (versus complete nucleosome eviction) appears to depend both on transcription rate (Bintu et al., 2011) and on polymerase density (Kulaeva et al., 2010).

Together, these studies show a strong influence of RNA polymerase, either alone or in conjunction with associated remodelers or chaperones, on nucleosome positioning. In particular, RNA polymerase broadly subverts intrinsic sequence preferences for nucleosomes, and furthermore appears to play a perhaps counterintuitive role in directional nucleosome movement upstream against the direction of transcription.

Integrating *cis* and *trans*-acting factors in nucleosome positioning

In the above sections, DNA flexibility and *trans*-acting factors were treated as independent contributors to nucleosome positioning. This is of course an oversimplification, as ATP-dependent remodelers often exhibit preferential positioning or nucleosome eviction specifically at pro- or anti-nucleosomal sequence elements. As noted above, in yeast the Isw2 ATP-dependent remodeler acts preferentially to position nucleosomes over intrinsically unfavorable sequences typically found at NDRs, thereby generally acting to shift nucleosomes over promoters (Whitehouse and Tsukiyama, 2006; Whitehouse et al., 2007). In contrast, SWI/SNF appears primarily to reinforce sequence effects

on nucleosome occupancy, as *snf2*Δ mutant yeast preferentially gain nucleosomes over intrinsically nucleosome-disfavoring promoters (Tolkunov et al., 2011). A multitude of similar cases can be found in the literature – the positioning of a nucleosome at a given sequence is influenced by the sequence's intrinsic preferences, but these preferences can be reinforced or subverted by everything from cytosine methylation to TF binding to transcription level. Thus, while many broad mechanistic themes have become apparent that play key roles in genomic nucleosome positioning, we feel it will be a long time before anyone has the ability to analyze any given 1 kilobase sequence and predict precisely where nucleosomes will assemble in a given organism.

Consequences of nucleosome positioning

It has long been apparent that the tight complex between histones and DNA would affect DNA accessibility, and this has motivated the general notion that nucleosomes primarily inhibit DNA-templated processes by blocking access to relevant factors. This is of course broadly true, but over the decades exceptions to this rule have emerged, and moreover the basic fact that nucleosomes interfere with DNA-binding proteins turns out to have endlessly complex implications for gene regulation, depending on the detailed layout of DNA-binding sites relative to nucleosome positions. Overall, while the rules underlying the establishment of nucleosome positioning are increasingly understood,

understanding the consequences of nucleosome positioning for cellular processes remains a bit of a dark art.

Nucleosomal repression of gene expression

As noted above, nucleosomes broadly inhibit access of proteins such as transcription factors to their binding sites. The classic demonstration that nucleosomes generally inhibit transcription comes from the Grunstein lab – depleting histones *in vivo* (by shutting off transcription of a galactose-driven pair of histones) resulted in derepression of *PHO5* under normally repressive high phosphate growth conditions (Han and Grunstein, 1988; Han et al., 1988a). Similar results have been reported for many other yeast genes in histone shutoff experiments, a result that has since been extended to the whole genome (Gossett and Lieb, 2012; Wyrick et al., 1999), with over 2400 genes being reported as derepressed in response to H3/H4 depletion. Genes derepressed in response to histone depletion were significantly correlated with the set of genes with decreased promoter nucleosome occupancy, again consistent with the general idea that promoter nucleosomes inhibit transcription. An alternative approach to tuning overall nucleosome occupancy at a specific promoter came from classic studies from Iyer and Struhl, who showed that increasing PolyA length, with resulting decreased nucleosome occupancy, at the *HIS3* promoter contributes to increased expression of this gene (Figure I.6A) (Iyer and Struhl, 1995). These studies are now being systematically extended by the Segal lab,

who have examined the expression of thousands of fluorescent reporter constructs driven by a variety of TFs, flanked by varying length PolyA elements (Figure I.6). Consistent with Iyer and Struhl's studies, inhibiting nucleosome binding to the promoter via strong PolyA elements or via addition of GRF binding sites contributed to increased reporter transcription (Sharon et al., 2012).

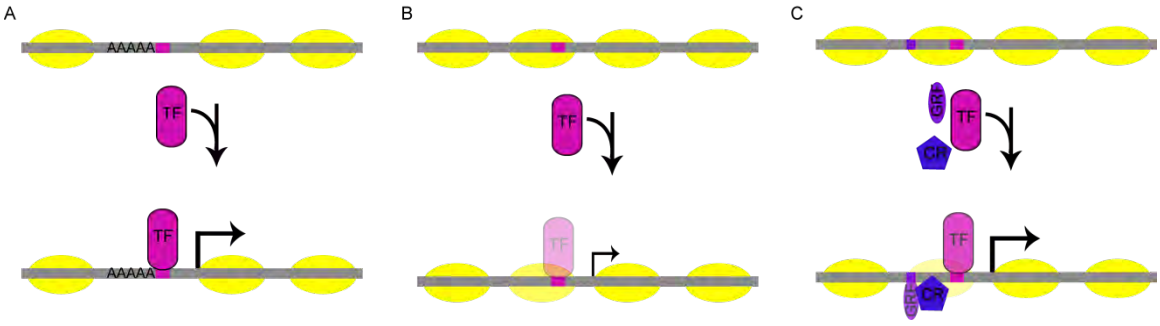


Figure I. 6: Promoter chromatin organization allows for tuning of gene expression.

Nucleosome depletion is generally conducive to gene expression and can be achieved by polyA elements programming an open promoter (A), or *trans*-acting factors abrogating histone binding (B) or moving nucleosomes (C). Genes with higher intrinsic nucleosome occupancy tend to be more reliant on *trans*-acting factors, such as general regulatory factor or chromatin remodellers, for higher levels of gene expression (B,C).

Because nucleosomes are generally repressive, differences in intrinsic nucleosome occupancy between promoters can have important consequences for gene regulation (Figure I.6). As noted above, in yeast highly-expressed "growth" genes have significantly AT-rich promoters, whereas "stress" genes are generally less AT-rich. As a result, stress gene promoters exhibit significantly greater nucleosome occupancy than do growth gene promoters. This results in dramatic regulatory differences between these gene classes – growth genes are generally unaffected by mutations in chromatin remodeling machinery, whereas stress genes are highly sensitive to a wide variety of chromatin-related mutants which are required to evict or destabilize nucleosomes in order to overcome nucleosomal repression of these promoters (Basehoar et al., 2004). Related results are seen in human cells, where for example genes induced upon TLR4 stimulation could be separated into Swi/Snf-dependent and –independent groups based on their promoters' intrinsic propensity to form nucleosomes (Ramirez-Carrozzi et al., 2009). As noted above, a curious evolutionary quirk is that the widely-expressed housekeeping genes in mammals, the apparent analog of yeast growth-related genes, are associated with GC-rich promoters.

The relatively high nucleosome occupancy of stress genes in yeast has many implications for gene regulation. Most trivially, these genes typically are expressed at lower levels than are growth genes. Perhaps more interestingly, these genes tend to exhibit more cell-to-cell variability in expression, or noise, than do growth genes – this will be discussed below. Additionally, inherently

nucleosome-repressed promoters provide for highly tunable regulation due to the intrinsic complexity of chromatin-based regulation (Figure I.6B,C). For instance, because many chromatin regulators are regulated by second messengers or key metabolites – including alpha ketoglutarate, acetyl-CoA, S-adenosylmethionine, NAD, and others – intrinsically nucleosome-occupied promoters are therefore potentially far more likely to respond to changes in cellular metabolism. Indeed, stress genes tend to be highly environmentally-responsive in yeast, although there are many other potential explanations for such behavior, including the obvious one that the TFs that drive these promoters are subject to environmental control.

Nucleosomal effects on RNA polymerase elongation

In addition to nucleosomal repression of transcription initiation *in vivo* by obstructing binding sites for transcription factors, nucleosomes can also present strong barriers for polymerase elongation – a template with a positioned nucleosome is inefficiently transcribed by RNA Polymerase II *in vitro*, even under high salt conditions (Bondarenko et al., 2006; LeRoy et al., 1998; Orphanides et al., 1998). RNA Polymerase II appears to pause about 45bp into the nucleosome, at the H3/H4 tetramer (Bondarenko et al., 2006), likely reflecting a requirement for DNA to stochastically unwrap from the histones to allow the polymerase to proceed further (Hodges et al., 2009). While nucleosomes may present a barrier to RNA Polymerases, they are not an insurmountable barrier.

For instance, running *in vitro* transcription assays with a second trailing polymerase, akin to what might occur during more rapid transcription *in vivo*, has been seen to aid the leading polymerase in its progress through the nucleosome, apparently thanks to the trailing polymerase inhibiting nucleosome-induced backtracking of the leading polymerase (Jin et al., 2010; Kulaeva et al., 2010). Furthermore, a number of “elongation factors”, including ATP-dependent chromatin remodelers and other chromatin regulators such as the H2A/H2B chaperone FACT play key roles in the alleviation of the nucleosomal barrier to RNA polymerase (LeRoy et al., 1998; Orphanides et al., 1998). Thus, a number of cofactors present *in vivo* can assist RNA polymerase in transiting chromatin templates. However, the extent to which chromatin elongation factors represent uniform constituents of elongating RNA polymerase, or are used in a promoter-driven manner that differs between genesets, remains an active area of investigation.

Nucleosomal activation of gene expression

While genome-wide analyses reveal thousands of transcripts in yeast that are de-repressed in response to histone depletion as expected, hundreds of transcripts surprisingly *decrease* expression upon histone depletion (Gossett and Lieb, 2012; Wyrick et al., 1999). This counterintuitive result points towards the underappreciated roles for nucleosomes in gene activation. A number of mechanisms exist by which nucleosomes can enhance gene transcription.

First, non-coding transcription is widespread throughout the genome and has been associated with suppression of coding genes (Martens et al., 2004); therefore, nucleosomal repression of inhibitory non-coding transcripts can result in gene activation. Second, several cases have been described in which nucleosomes contribute positively to transcription by taking up DNA between two DNA binding sites to bring the regulatory elements into physical proximity. For example, at the human U6 promoter, two DNA-binding sites are located ~150 bp away from one another. In the absence of a nucleosome, the TFs that bind these sites each bind too weakly to effectively activate transcription, and the distance separating the motifs is too great for significant cooperation between the TFs. However, upon incorporation into a nucleosome with the resulting compaction of the intervening DNA, the two TF binding motifs are brought into much greater spatial proximity, allowing the TFs to bind cooperatively (Stünkel et al., 1997). Similar results were observed at the fly *hsp26* promoter (Lu et al., 1995). Additionally, while most transcription factors bind poorly to nucleosomal DNA, some are able to bind DNA on the surface of the nucleosome (McPherson et al., 1993) or compete with histones for binding to DNA (Workman and Kingston, 1992). Finally, a number of sequence-specific DNA-binding proteins also carry additional domains such as bromodomains or PHD fingers that are known to bind to modified histones. Thus, for a subset of transcription activators a chromatin context may enhance binding to a target promoter relative to a long naked stretch of DNA, by enabling bivalent binding interactions between the TF and both the

target motif and a nearby histone mark at the promoter in question. While these and other mechanisms have been explored in specific cases, a systematic understanding of nucleosome-dependent activation of genes lags far behind the converse case of chromatin-mediated gene repression.

Nucleosomal effects on gene regulatory logic and expression noise

The general idea that nucleosomes interfere with DNA binding by other proteins is a simple concept with endlessly complicated implications – because protein factors carry out the majority of the regulatory work for DNA-templated processes, nucleosomes can essentially have any arbitrary effect on gene regulation or origin firing, etc., based simply on which proteins they are in the way of. Noted here a few examples of paradigms in which nucleosomes have more complicated consequences for transcription than simple gene repression/activation.

A particularly interesting example of nucleosomal control of gene regulatory logic comes from the human β-interferon promoter, which is regulated by multiple transcription factors including NF-κB, IRF1/3/7, and ATF2. Here, a nucleosome positioned over the TATA box prevents transcriptional induction in response to activation of any one of the transcription factors (in response to, for instance TNF-α or IFN-γ). However, upon viral infection, all three transcription factors are activated, the nucleosome shifts downstream, and IFN-β is transcribed. In an elegant approach, Lomvardas and Thanos designed an

artificial IFN-β promoter in which a strong nucleosome positioning sequence was placed at the normal post-induction location (Lomvardas and Thanos, 2002). As a result, activation of any of the three transcription factors was sufficient to activate transcription. Thus, a 40 bp shift in nucleosome location can change a promoter from an AND gate to an OR gate.

Nucleosomes can also play roles in filtering signal strength and signaling kinetics. Nucleosomes are not fixed entities – histone proteins can be evicted in response to signaling cascades (Almer et al., 1986), and also exhibit dynamic eviction and replacement under steady-state conditions (Dion et al., 2007; Rufiange et al., 2007). As a result, nucleosome-occluded regulatory information can contribute to gene regulation in complex ways that are predicted to depend very strongly on the dynamics of promoter "opening/closing." For example, O'Shea and colleagues found that DNA-binding sites for the transcription factor Pho4 have different effects on *PHO5* transcription depending on whether they are exposed or nucleosome-occluded: exposed Pho4 sites contributed to the level of phosphate starvation required for *PHO5* induction, but after this threshold was reached and promoter nucleosomes were evicted, the sum total of all Pho4 binding sites contributed the extent of mRNA production (Lam et al., 2008). In this case, whatever promoter dynamics exist under uninduced conditions apparently do not expose Pho4 binding sites for long enough to have any regulatory impact, as these nucleosome-occluded TF binding sites solely contribute to mRNA production once exposed by nucleosome eviction. A similar

phenomenon has been reported for the yeast *CLN2* promoter, where multiple General Regulatory Factors contribute to a constitutive NDR covering the binding site for regulator SBF (Bai et al., 2010, 2011). Here, eliminating the NDR by mutating GRF binding sites, or moving the SBF binding site to a nucleosome-occluded site, resulted in extensive cell to cell variability in *CLN2* expression. In contrast, *CLN2* driven by an accessible SBF binding site occurred uniformly every cell cycle. Most interesting in these studies was the observation that loss of the NDR had no effect on the level of *CLN2* in individual expressing cells, but instead altered only the fraction of cells expressing this reporter. Thus, similar to Lam and O'Shea's studies on *PHO5*, nucleosome-occluded regulatory information becomes functional in those cells in which this information is exposed by nucleosome eviction.

Understanding quantitatively the effects of nucleosome-occluded regulatory information will require integrating information about histone exchange dynamics with DNA sequence information and TF dynamics. In general, promoters with high steady-state nucleosome occupancy are highly responsive to chromatin mutants, and exhibit high levels of cell to cell variation ("noise") in gene expression (Field et al., 2008; Newman et al., 2006; Raser and O'Shea, 2004; Tirosh and Barkai, 2008). While much of this may be driven by the high enrichment of TATA boxes at such genes (Basehoar et al., 2004) and the resulting slow off rate for bound TBP, modeling and experimental studies also point to a major influence of chromatin state transitions from "open" to "closed"

on gene expression variability (Boeger et al., 2008; Raser and O'Shea, 2004). Chromatin dynamics at promoters have more subtle regulatory effects than simply influencing cell to cell variation in expression, as they have also been implicated in how promoters filter information from transcription factor dynamics. Here, combined modeling and experimental studies suggest that promoters with slow open/closed transition kinetics are generally less responsive to transcription factor activity that is frequency or duration-modulated, whereas both slow and fast promoters are similarly regulated by amplitude-modulated transcription factors (Hao and O'Shea, 2012).

Chromatin structure and regulatory evolution

Early comparative genomics studies in fungi identified a strong connection between genes that changed regulatory behavior between species, and promoter chromatin architecture. In general, genes that exhibit expression "volatility" between species – genes that are highly-expressed under conditions of rapid growth in one species, but poorly-expressed under similar conditions in another – are the "stress" class of genes with high promoter nucleosome occupancy (Tirosh et al., 2006). Moreover, changes in gene regulatory program can often be linked to changes in nucleosome occupancy. One of the first-described and most dramatic examples of this behavior is found in the mitochondrial ribosomal protein (MRP)-encoding genes in fungi – Ascomycete species that diverged in this phylogeny prior to a whole-genome duplication (WGD) event (such as *C.*

albicans) typically harvest energy from carbohydrates via respiration during rapid growth, whereas species diverging post-WGD exhibit rapid fermentation-dependent growth, and then switch to respiration after all fermentable carbon has been depleted (Conant and Wolfe, 2007). Genes related to respiration, such as mitochondrial ribosomal protein-encoding genes (and many other related genesets), are regulated with other "growth" genes in pre-WGD species, but are regulated with "stress" genes in post-WGD species (Ihmels et al., 2005). This change in regulatory strategy is linked to sequence effects on nucleosome occupancy, as pre-WGD species carry more antinucleosomal PolyA tracts at these genes and thus "program" more open chromatin, while PolyA tracts are lost at these genes in post-WGD species (Field et al., 2009; Tsankov et al., 2010). Similar changes occur at other genesets on this phylogeny, as for example splicing-related genes in *Y. lipolytica* are associated with stronger PolyA sequences and more highly-expressed during midlog growth than they are in relatively intron-depleted species elsewhere in the phylogeny (Tsankov et al., 2010).

Over shorter evolutionary timescales, both *cis* and *trans*-acting sequence changes affect chromatin structure and gene regulation. Seminal studies by Brem and Kruglyak analyzed segregants of a cross between closely related *S. cerevisiae* strains to identify *cis*- and *trans*- effects of polymorphisms on gene expression levels (Brem et al., 2002). Subsequent analysis mapped many of these *trans*-acting polymorphisms to chromatin regulators (Lee et al., 2006). In

contrast, Tirosh and Barkai used hybrid diploids formed from closely-related *S. cerevisiae* and *S. paradoxus* to show that most gene expression divergence was driven by *cis*-acting sequence differences. Much of this *cis*-acting sequence divergence ended up affecting the strength of antinucleosomal sequences at promoters (Tirosh et al., 2009, 2010). Thus, while gene regulatory divergence in the short term BY/RM system were driven by *trans*-regulators, and longer-term gene expression divergence was linked to *cis*-acting sequence changes, both of these effects appear to be mediated via chromatin changes.

Comparitive genomics studies generally find that evolutionarily labile genes are associated with highly nucleosome-occupied promoters, but of course it is quite plausible that this has no causal role in gene expression changes over evolutionary time, instead reflecting either relaxed selective constraints, or even positive selection, on expression levels of “stress” genes relative to the generally highly-conserved “growth” genes. However, several laboratory evolution studies have shown that the chromatin state of a promoter indeed affects the available paths to altered gene expression. Verstrepen and colleagues showed that by selecting for increased expression of a reporter gene driven by a PolyA-containing promoter, they could obtain yeast strains with increased PolyA lengths (Vinces et al., 2009). More broadly, Barkai and colleagues carried out selection experiments on yeast strains engineered to drive GFP from a variety of different promoters. Increased expression could be selected for GFPs driven both by relatively nucleosome-free and nucleosome-occupied promoters (Rosin et al.,

2012). However, strains selected for increased expression from nucleosome-free promoters typically exhibited gene duplications, and upon removal of selection pressure duplicated regions were rapidly lost. In contrast, nucleosome-occluded promoters could easily be selected to drive higher GFP, and the mutations that altered GFP expression were primarily unlinked to the reporter (eg in *trans*). Moreover, analysis of unselected mutations on a variety of reporters confirmed that nucleosome-occupied promoters were broadly more sensitive to mutations than were nucleosome-depleted promoters (Hornung et al., 2012). These results strongly argue that the chromatin structure of promoters contributes to the evolvability of new gene regulatory programs.

In the above cases, much of the impact of chromatin architecture on evolution of gene regulation can be ascribed to the impact of mutational “target size” for highly nucleosome-occupied promoters, whose expression is affected by many more chromatin regulators. However, nucleosomes can also influence the spectrum of mutations at a given sequence, as nucleosomal DNA is less available to DNA repair enzymes than is linker DNA. This is observed in analyses of genetic variation within a species or between closely-related species, as for example in Medaka different types of sequence polymorphisms are observed to occur at genomic loci found in nucleosomes (nucleotide substitutions) vs. in linkers (indels) (Sasaki et al., 2009). These and similar results suggest that nucleosomes influence the course of sequence evolution over evolutionary time, thus casting a sequence “shadow” on the genome.

CHAPTER II: High-resolution nucleosome mapping reveals transcription-dependent promoter packaging.

Abstract

Genome-wide mapping of nucleosomes has revealed a great deal about the relationships between chromatin structure and control of gene expression, and has led to mechanistic hypotheses regarding the rules by which chromatin structure is established. High-throughput sequencing has recently become the technology of choice for chromatin mapping studies, yet analysis of these experiments is still in its infancy. Here, we introduce a pipeline for analyzing deep sequencing maps of chromatin structure and apply it to data from *S. cerevisiae*. We analyze digestion series where nucleosomes are isolated from under- and over-digested chromatin. We find that certain classes of nucleosomes are unusually susceptible or resistant to overdigestion, with promoter nucleosomes easily digested and mid-coding region nucleosomes being quite stable. We find evidence for highly sensitive nucleosomes located within "nucleosome-free regions," suggesting that these regions are not always completely naked but instead are likely associated with easily-digested nucleosomes. Finally, since RNA polymerase is the dominant energy-consuming machine that operates on the chromatin template, we analyze changes in chromatin structure when RNA polymerase is inactivated via a temperature-sensitive mutation. We find evidence that RNA polymerase plays a role in nucleosome eviction at promoters, and is

also responsible for retrograde shifts in nucleosomes during transcription. Loss of RNA polymerase results in a relaxation of chromatin structure to more closely match in vitro nucleosome positioning preferences. Together, these results provide analytical tools and experimental guidance for nucleosome mapping experiments, and help disentangle the interlinked processes of transcription and chromatin packaging.

Introduction

Eukaryotic DNA is packaged in nucleosomes, composed of 147bp of DNA wrapped ~1.7 turns around an octamer of histone proteins (Kornberg and Lorch 1999; Luger et al. 1997). Nucleosomes influence the expression of a huge fraction of yeast genes (Wyrick et al., 1999), and the precise positioning of nucleosomes relative to underlying DNA, controls access to protein binding sites and thereby affects regulatory programs (Lam et al. 2008; Lomvardas and Thanos 2002; Radman-Livaja and Rando 2009; Stunkel et al. 1997). In yeast, transcription start sites (TSSs) are generally positioned just within the +1 nucleosome, downstream of a nucleosome-depleted region generally referred to as the nucleosome-free region (NFR) (Albert et al. 2007; Mavrich et al. 2008a; Yuan et al. 2005). Nucleosome positioning is different between promoter types; TATA-less promoters tend to be “housekeeping” genes and are characterized by a canonical NFR upstream of their TSS; conversely, TATA-containing genes tend to be stress-responsive, noisily expressed, are more responsive to genetic

perturbation of chromatin remodeling complexes, and their promoters are often at least partly occupied by nucleosomes (Basehoar et al. 2004; Choi and Kim 2009; Field et al. 2008; Ioshikhes et al. 2006; Newman et al. 2006; Tirosh and Barkai 2008).

While nucleosomes can package almost any sequence, work over the past few decades has established sequence rules that influence this packaging. Most importantly, rigid polyA tracts in DNA are unfavorable for nucleosome assembly, and can direct formation of NFRs in vitro (Drew and Travers 1985; Ioshikhes et al. 2006; Iyer and Struhl 1995; Kaplan et al. 2008; Kunkel and Martinson 1981; Segal and Widom 2009; Sekinger et al. 2005; Yuan and Liu 2008; Yuan et al. 2005). This was recently confirmed globally in an in vitro reconstitution study, emphasizing a major role for nucleosome-excluding sequences in “programming” promoter architecture (Kaplan et al. 2008; Zhang et al. 2009). Much of the remainder of in vivo chromatin structure is proposed to result from “statistical positioning,” the idea that packing as many nucleosomes as possible in a short stretch of the genome will result in positioned nucleosomes (Kornberg and Stryer 1988; Mavrich et al. 2008a; Yuan et al. 2005). While some aspects of yeast chromatin structure can in principle be predicted based on sequence and packing rules, a number of protein machines, such as the ATP-dependent chromatin remodeling complexes, regulate transcription by moving nucleosomes (Clapier and Cairns 2009; Workman and Kingston 1998). Perhaps the most widespread chromatin-perturbing complex is RNA polymerase, whose

passage disrupts histone-DNA contacts, and at very high transcription rates results in nucleosome eviction (Bondarenko et al. 2006; Dion et al. 2007; Field et al. 2008; Lee et al. 2004; Schwabish and Struhl 2004; Studitsky et al. 1994; Studitsky et al. 1997).

The advent of high-throughput sequencing technology has enabled rapid genome-wide analysis of nucleosome positioning at high resolution, and has been used to map nucleosomes in yeast, worms, flies, medaka, and humans (Field et al. 2008; Mavrich et al. 2008a; Mavrich et al. 2008b; Sasaki et al. 2009; Schones et al. 2008; Shivaswamy et al. 2008; Valouev et al. 2008). Yet, analytical and experimental methods for deep sequencing analysis of chromatin are still in their infancy and mostly focus on averaged nucleosome occupancy levels at genomic loci, effectively transforming deep sequencing data into an analog of tiling array data.

Here, we describe a novel analytical method for identifying nucleosome positions from Illumina sequencing data, and automatically estimate nucleosome position, occupancy, and length from *S. cerevisiae* data. Using parameters extracted from these nucleosome calls allows us to identify groups of functionally-related genes with significantly high or low values of a given parameter. Experimentally, we examine the impact of nuclease titration levels on this assay and identify properties of yeast chromatin that are influenced by the level of digestion, as well as invariant properties. Finally, we explore the role of

RNA polymerase in chromatin structure through the use of a temperature-sensitive mutant in the gene encoding the large subunit of RNA polymerase II, *RPO21* (also known as *RPB1*). Comparing nucleosome positions and properties before and after Pol II inactivation we confirm the role of RNA polymerase in nucleosome eviction at promoters, and find a surprising role in retrograde movement of nucleosomes over genes. By quantitatively analyzing changes in nucleosome positioning after Pol II shutoff we find that different classes of genes are subject to distinct perturbations by RNA Polymerase. Finally, we confirm a role for RNA Polymerase in perturbing nucleosomes from their thermodynamically-favored positions.

Results

Identifying Nucleosome Positions by Template Filtering

The application of deep sequencing to mononucleosomal DNA results in million of reads from both ends of the mononucleosomal DNA segments. Published methods for calling nucleosome positions from Illumina data typically involve extending each single-end short sequenced read to the expected segment length of ~140 bp, and then examining the coverage of different genomic loci by the accumulated extended segments. These methods clearly highlight nucleosome-depleted vs. nucleosome occupied regions, for example in averaged gene alignments (Albert et al. 2007; Field et al. 2008; Shivaswamy et al. 2008). More elaborate approaches identify nucleosome positions by

identifying the center of these inferred segments and estimating the occupancy at different center locations (Shivaswamy et al. 2008). These methods do not account for the possibility that nucleosomes might occupy different positions in subpopulations of cells, and assume uniform nucleosome lengths (ie no variability in digestion level at different nucleosomes).

To overcome these issues, we developed a method for calling nucleosome positions, occupancy and length, using template filtering (Turin 1960). This method is based on the observation that sequencing the ends of a nucleosome will result in an expected pattern of offset Forward and Reverse strand reads at the two ends (Figure II.1A). Due to variability in exact nucleosome position from cell to cell, and variability in the extent of MNase digestion at each end of a nucleosome, the peak of reads at each nucleosome end will form a distribution of variable width. Our method is based on identifying occupancy templates of Forward and Reverse reads that are typical of nucleosomes. The method then uses a fast procedure to identify locations where the Forward and Reverse read distributions correlate with a series of model templates (Figure II.1B, Supplementary Figure II.1). The method examines a range of distances between Forward and Reverse templates to capture over- and under-digestion of the ~147bp nucleosomal DNA, and determines nucleosome length by choosing an F-to-R offset that maximizes correlation to F and R templates (Figure II.1C,D).

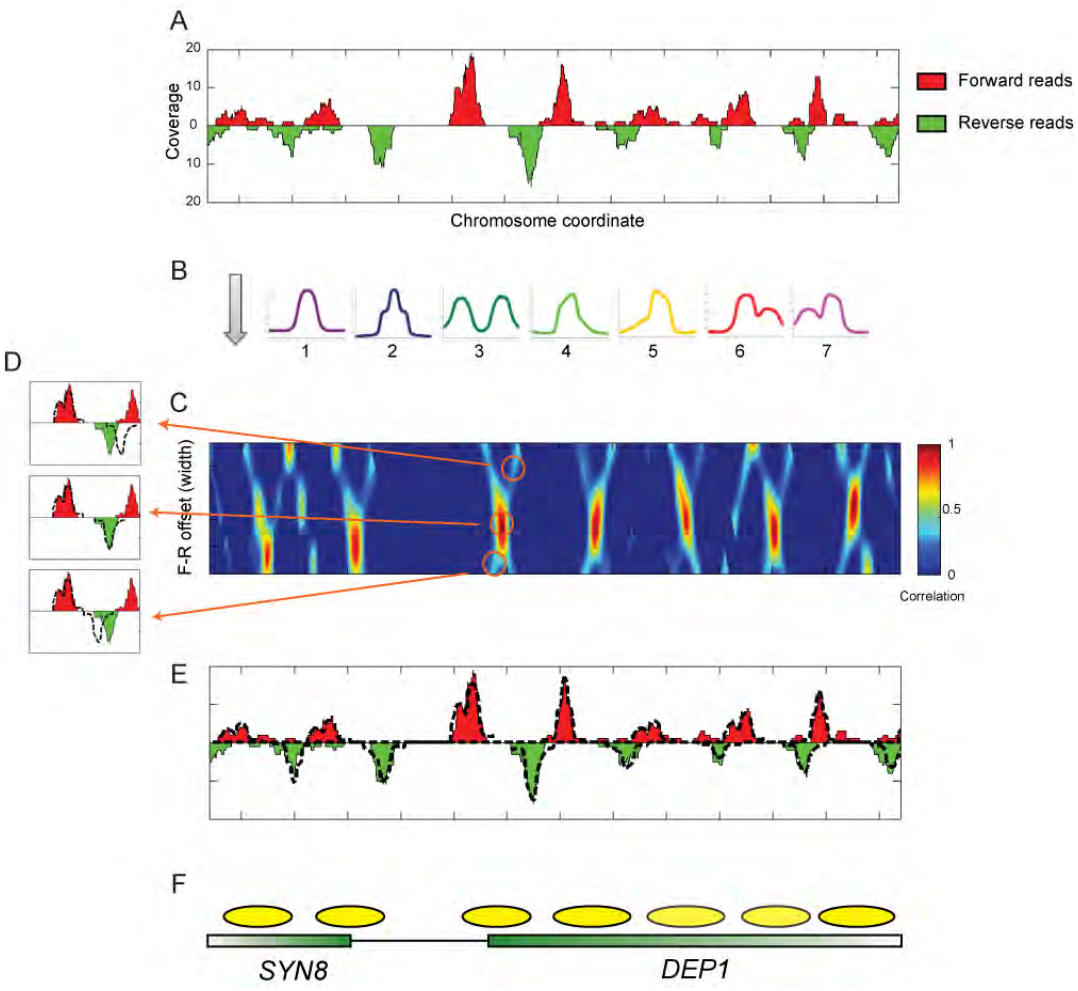


Figure II.1: Template Filtering overview

A) Deep sequencing data for a typical stretch of the yeast genome. Coverage by forward strand sequencing reads are shown as red peaks, whereas coverage by reverse strand sequencing reads are shown as inverted green peaks.

B) Templates. Forward and Reverse strand read distributions are cross-correlated with each of the seven templates shown.

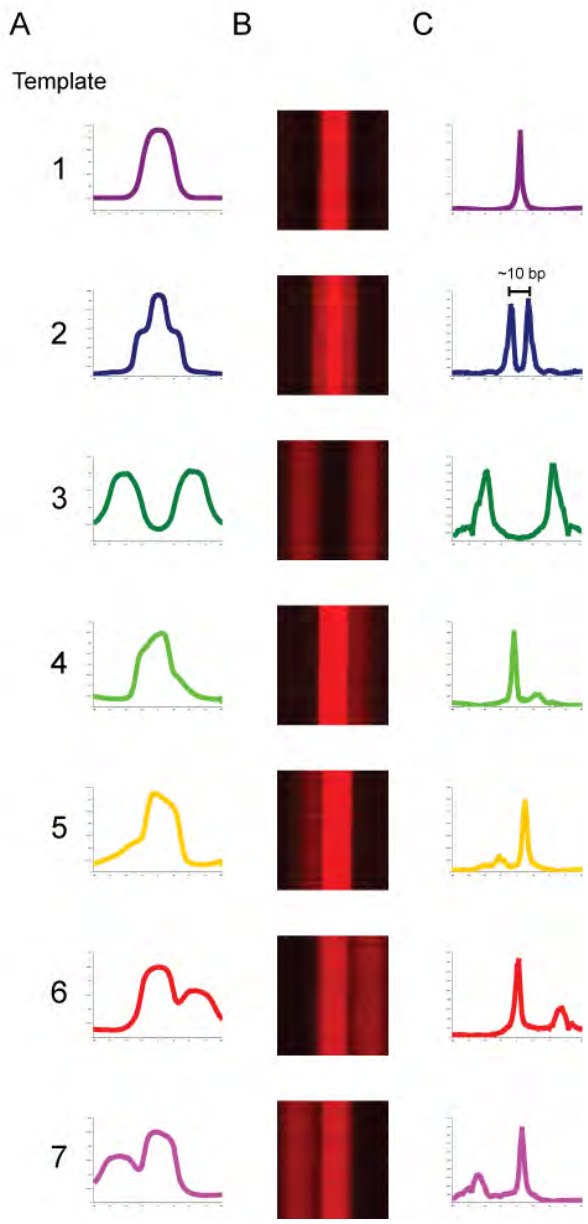
C) Correlation coefficient heat map of template 1 for Forward and Reverse templates, at varying center positions (x-axis) and distances (y-axis).

D) Examples of templates spaced too far apart (top), at the optimal distance (middle), or too close together (bottom). Dotted lines indicate template outlines being compared to the underlying data.

E) Read distributions explained by the optimal template matches are shown as dotted lines for the region from part A).

F) Schematic of nucleosome calls and underlying gene annotations.

To further elaborate the method, we aimed to identify the common read distributions from the data. In other words, what are the typical distributions of end reads for the nucleosomes in an experimental dataset? We first applied our method using a Gaussian-shaped template to create a preliminary map of nucleosome positions. We next examined read distributions at the ends of these initial nucleosome predictions, clustered read patterns to identify typical behaviors, and selected seven representative templates for use in our scans (Figure II.1B, Supplementary Figure II.1). Six of the templates exhibited several peaks, potentially indicating nucleosomes with variable ends resulting from subpopulations and/or variable nuclease digestion. For example, nucleosomes with ends that match template 2 have sets of ends separated by ten base pairs (Supplementary Figure II.1), suggesting that the first stage in overdigestion of nucleosomes by micrococcal nuclease is to cut one helical turn further into the nucleosomal DNA.



Supplementary Figure II.1: Templates used for filtering

A) The seven templates used for pattern-matching.

B) Read distributions for nucleosome matching each template. All nucleosome ends associated with a given template are represented in a heatmap, with each row representing a nucleosome and red bars indicating the 36 nt sequenced reads.

C) For all nucleosome ends matching the indicated template, histograms are displayed for all sequence reads associated with the template.

Our method fits the sequencing data to the best-correlated templates with varying distances between F and R templates. Local maxima in the correlation spectrum (Figure II.1C) are identified as potential nucleosome calls (Figure II.1E,F). To assemble the final set of nucleosomes, we use a greedy approach to choose the best-correlating template and distance per nucleosome (Methods). Occupancy is determined for each nucleosome by the number of reads contributing to a given nucleosome call. To account for different sequencing yields we normalize nucleosome occupancies to mean value of 1.

To evaluate the quality of our method's nucleosome calls, we used the nucleosome positions, templates, and occupancy calls to re-generate the sequenced reads which account for our called nucleosomes. This simulated dataset is sampled with the same number of reads as the original sequencing run. These simulated reads closely match the original sequencing data, and the small difference (residual) between the measured and reconstructed data indicate that our nucleosome calls account for 88% of the sequencing reads, indicating that the small number of extracted parameters capture the majority of the experimental sequencing data (Supplementary Figure II.2).

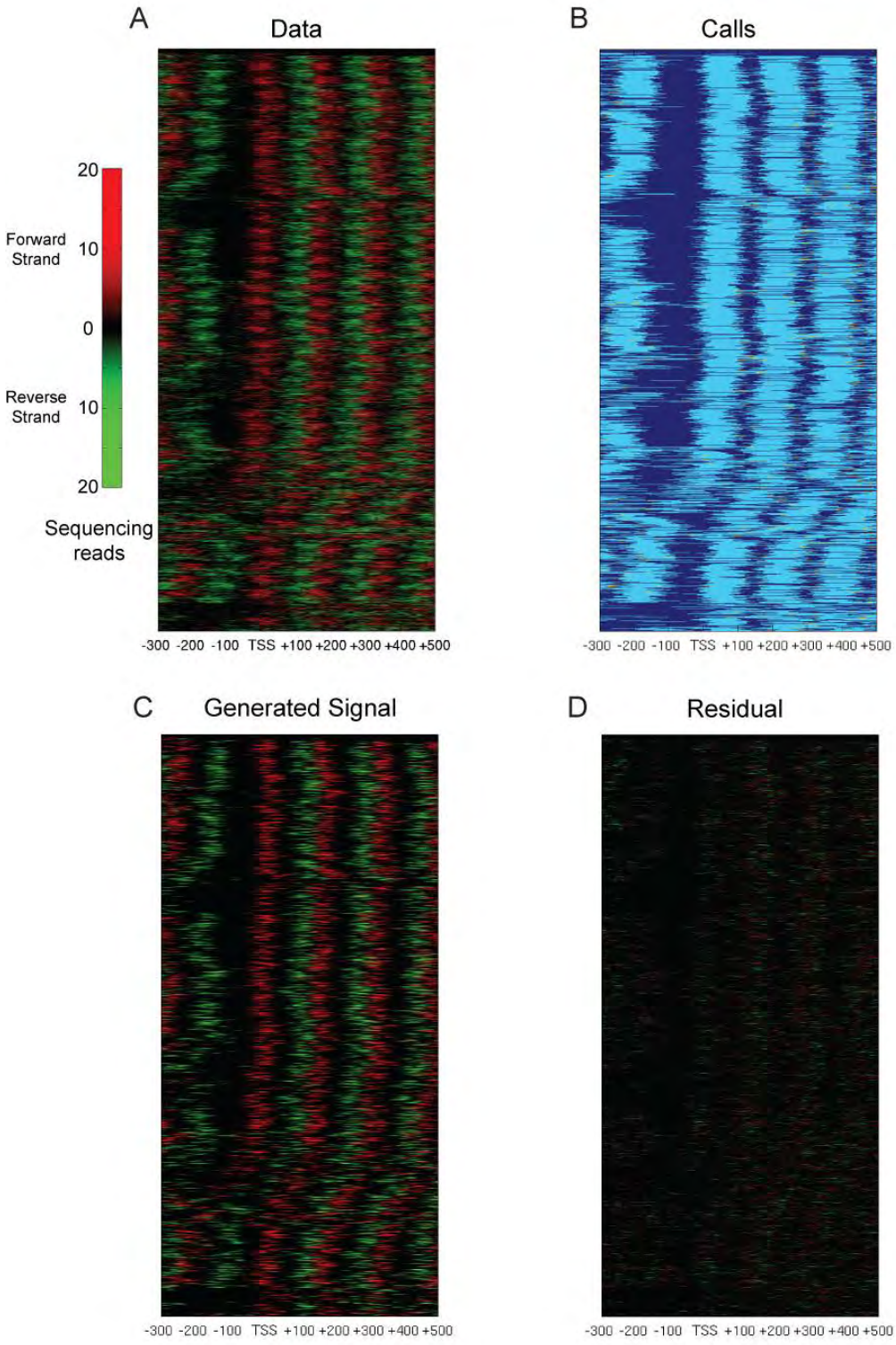
Supplementary Figure II.2: Nucleosome calls account for the majority of the sequencing data

A) Occupancy map of forward and reverse strand reads for all yeast ORFs aligned by TSS. Forward strand reads are shown as red bars, reverse strand reads are shown in green. Gene order for parts B-D is the same as for this panel.

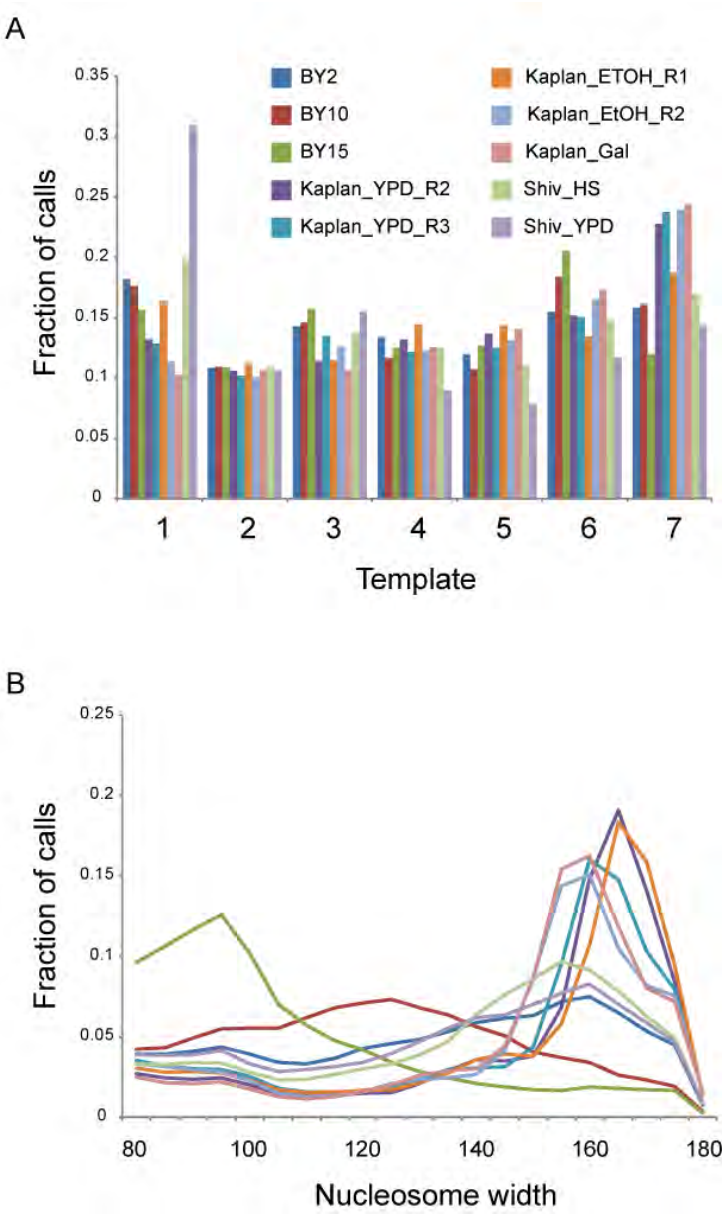
B) Nucleosome calls generated by our method are shown in blue. Small overlaping areas are shown with yellow

C) Simulated data regenerated from our nucleosome calls only (using selected templates and occupancy), shown as in A).

D) Generated data was subtracted from real data. Small difference (residual) between raw data and regenerated data demonstrates that our method accounts for 88% of raw data.


A
Data
B
Calls
20
Forward Strand
10
0
Reverse Strand
10
20
Sequencing reads
-300 -200 -100 TSS +100 +200 +300 +400 +500
-300 -200 -100 TSS +100 +200 +300 +400 +500
C
Generated Signal
D
Residual
-300 -200 -100 TSS +100 +200 +300 +400 +500
-300 -200 -100 TSS +100 +200 +300 +400 +500

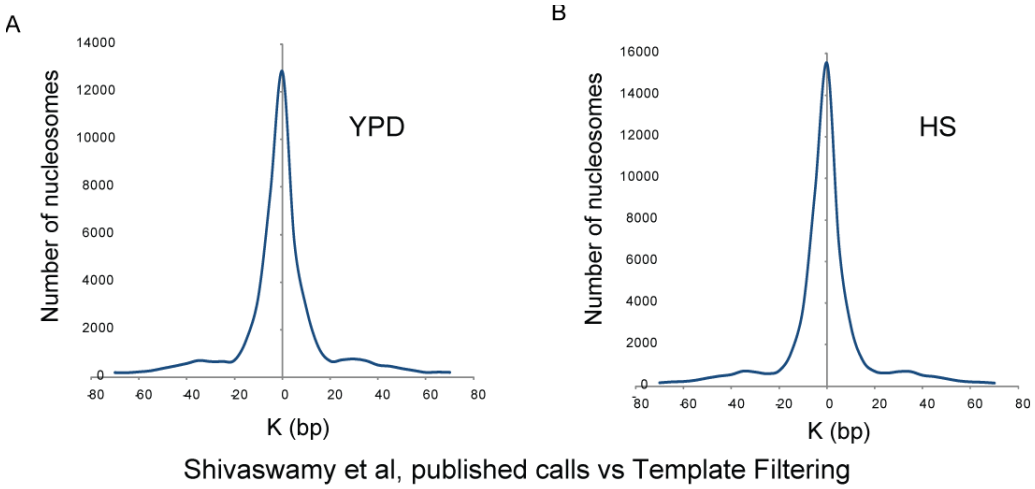
We also examined data from 7 additional previously-published deep sequencing datasets (Kaplan et al. 2008; Shivaswamy et al. 2008). We find that the same templates are common in all datasets (Supplementary Figure II.3A), confirming the generality of our approach. However, we did find that the occurrence frequency of different templates differed between datasets, which we ascribe to differences in numbers of reads as well as differences in digestion between different MNase preparations (Supplementary Figure II.3B, see below). Positions of nucleosome calls were generally concordant between datasets (Supplementary Figures II.4, II.5), with the major difference occurring in -1 nucleosome calls, which we also identify below as being dependent on digestion levels.

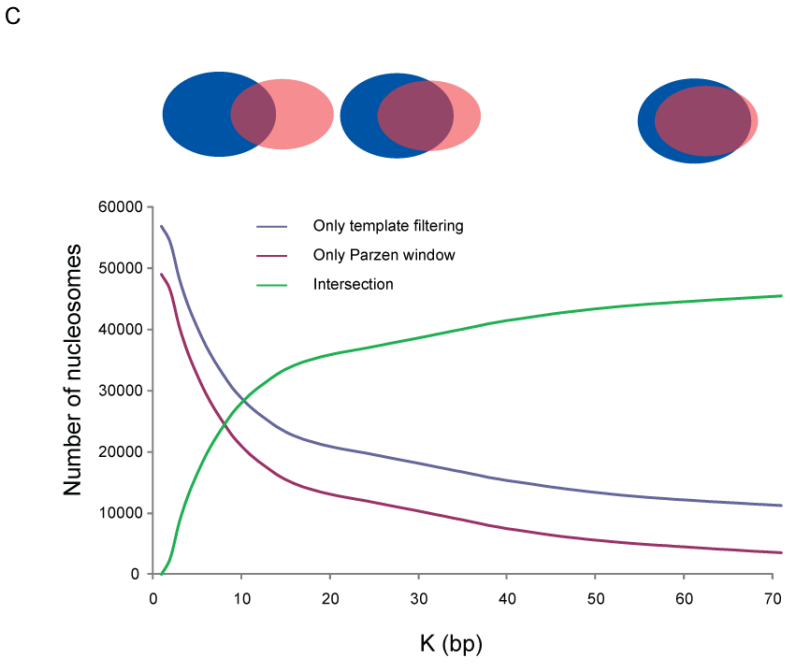


Supplementary Figure II.3: Template filtering applied to other Illumina datasets

A) Template filtering was applied to seven additional published nucleosome mapping datasets. Shown are frequencies of various templates for all ten (Seven published plus three from this study) sets of nucleosome calls.

B) Nucleosome width distributions from various datasets. Shown are histograms of nucleosome widths as determined by template filtering for the ten datasets above. Note that the two other studies both exhibit nucleosome width distributions corresponding to our “underdigested” chromatin.



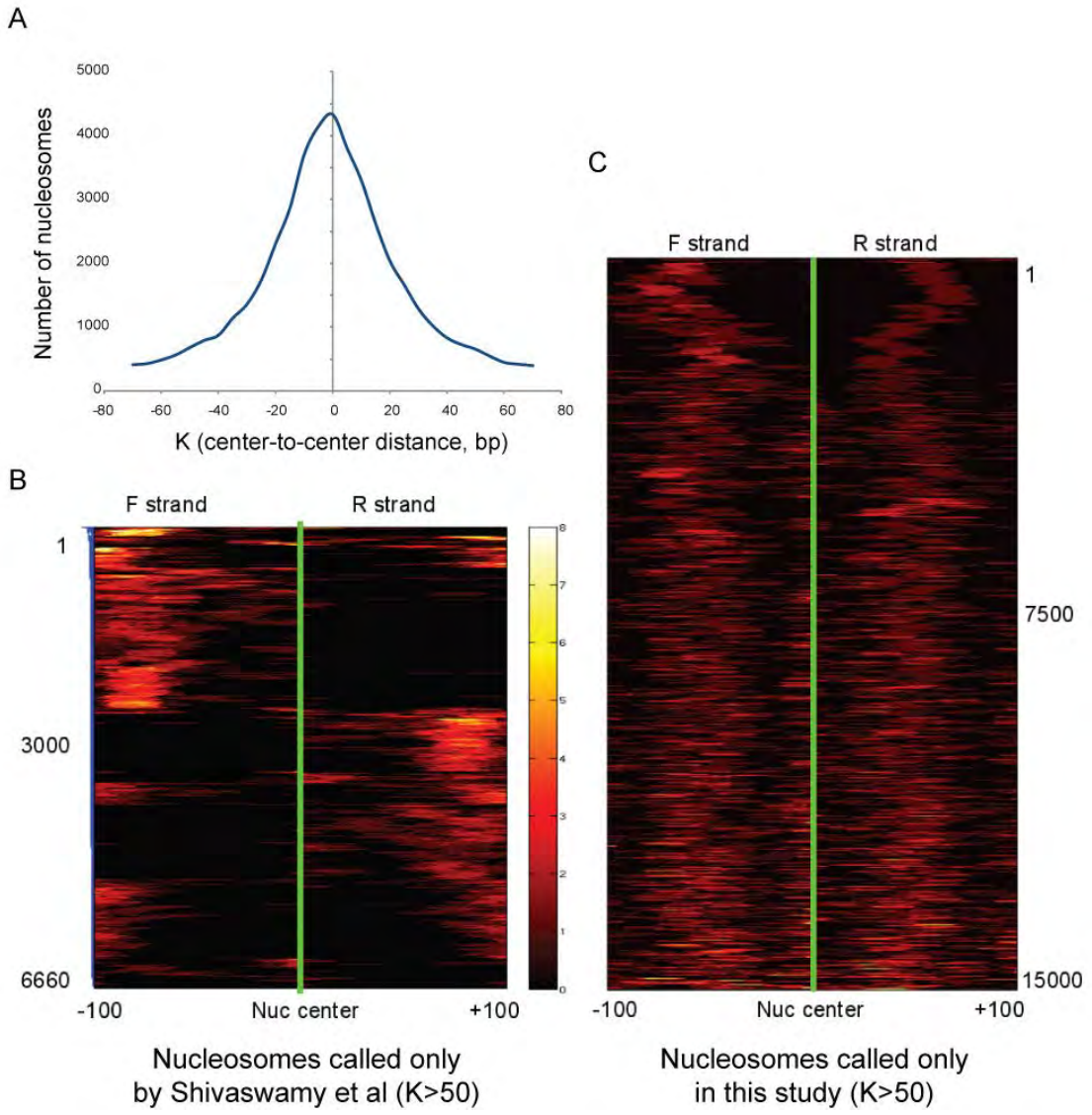


Supplementary Figure II. 4: Comparison of template filtering and Parzen window–based approach

A, B) Center-to-center distance histograms for nucleosome calls from Shivaswamy et al compared to template filtering calls using Shivaswamy data.

C) Curves showing nucleosomes called in this study vs. calls from Shivaswamy et al. For each center-to-center distance, calls unique to one study or the other are plotted, with green curve showing number of nucleosomes called with both methods with the indicated center-to-center distances. Venn diagrams above graph show intersection of calls at three values of K.

Finally, we compared our nucleosome calling method to an alternative calling algorithm based on identifying peaks in data where Forward and Reverse reads have been shifted a half-nucleosome width towards one another (Shivaswamy et al. 2008). Nucleosome calls were highly concordant for both methods (Supplementary Figure II.5A), particularly for clearly well-positioned nucleosomes (those that match Template 1 in our approach). Examination of regions where nucleosome calls differed revealed that our method fails in regions where only one end of a nucleosome generates reads, whereas our method better captures data from “fuzzy” regions such as mid-coding regions (Supplementary Figure II.5B,C), thereby better-capturing nucleosome occupancy over such regions.



Supplementary Figure II. 5: Differences between template filtering and Parzen window–based approach.

A) As in Supplementary Figure II.4A,B, but comparing nucleosome calls from this dataset to Shivaswamy et al calls on their dataset.

B) Heatmap of sequencing reads from nucleosomes called by Shivaswamy et al but not called using template filtering on Shivaswamy data. Note that most missed calls correspond to read distributions that only capture one nucleosome end. Note that this dataset is relatively undersequenced (~0.5 million reads), and this type of artifact is rarer in higher-coverage datasets.

C) Heatmap of sequencing reads from nucleosomes called using template filtering but not called by Shivaswamy et al. Note that most missed calls generally correspond to “fuzzy” stretches of chromatin.

By examining distributions of end reads, we found that nucleosomes at different locations vary in their digestion patterns. Over- and under-digestion of DNA can be due to many possible factors, including properties of the DNA sequence (e.g., sequence composition, bendability, etc.), and properties of the nucleosome (e.g., histone modification state). To investigate these two factors we tested whether the digestion template at nucleosome ends is associated with specific sequence properties and/or specific modification annotation. Consistent with previous reports of MNase sequence preference (Dingwall et al. 1981; Horz and Altenburger 1981), we find that different templates are associated with a clear sequence preference for location of A/T dinucleotides (Supplementary Figure II.6B). On the other hand, we also observed that multimodal templates such as template 3 were enriched at locations previously described as “fuzzy” or delocalized, such as mid-coding regions and over promoters of stress-responsive genes (Supplementary Figure II.6C,D). These results suggest that both nucleosomal subpopulations (ie delocalization) as well as sequence composition are significant sources of variability in patterns of nucleosome reads. While these two factors cannot be completely disentangled when using MNase to map nucleosomes, we note that the use of several templates provides an automated means for taking MNase sequence biases into account.

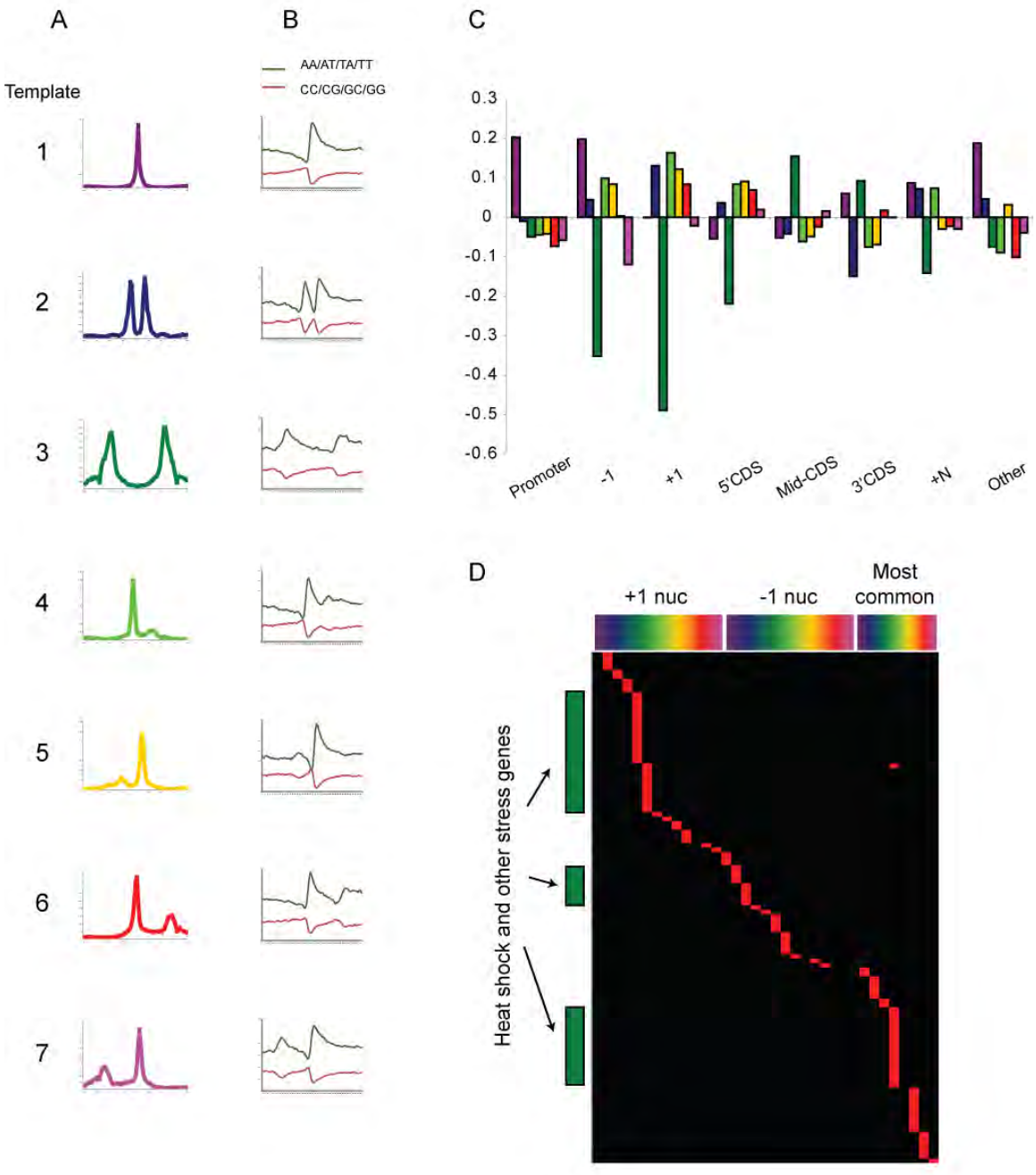
Supplementary Figure II. 6: Different templates capture MNase biases and delocalized nucleosomes

A) Templates start distribution, as in Figure II.1C. All x axes are as for Template 1.

B) Sequences of nucleosome ends associated with each template exhibit MNase-related biases. Nucleosomes of each template were aligned by their middle and dinuclotide frequency at each position were collected. Frequency of AA/AT/TA/TT dinucleotides at the sequence start are shown in green, and of GG/GC/CG/CC dinucleotides in red. Note strong bias for AT-rich dinucleotides at MNase cleavage sites – over 70% of all cut sites were after AA/AT/TA/TT. All x and y axis labels are identical to Template 1, so are excluded to reduce clutter.

C) Association of nucleosomes at varying genomic locations with different templates. Note that Template 3, representing widely-separated MNase cleavage sites, is enriched in mid-coding regions.

D) Enrichment of different templates at the -1, +1, and coding region average nucleosomes for various genesets (as in Figure II.2B). For +1 and -1 nucleosomes, two columns are show for each template, the first being use of the relevant template for the upstream nucleosome end and the second for the downstream end. Color bar on the top matches template colors from A, and green lines to the left show enrichment of template 3 in heat shock and stress genes.


A
B
C
D
Template
1
2
3
4
5
6
7
AA/AT/TA/TT
CC/CG/GC/GG
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
Promoter
-1
+1
5'CDS
Mid-CDS
3'CDS
+N
Other
+1 nuc
-1 nuc
Most common
Heat shock and other stress genes

Nucleosome Positioning in Growing Yeast

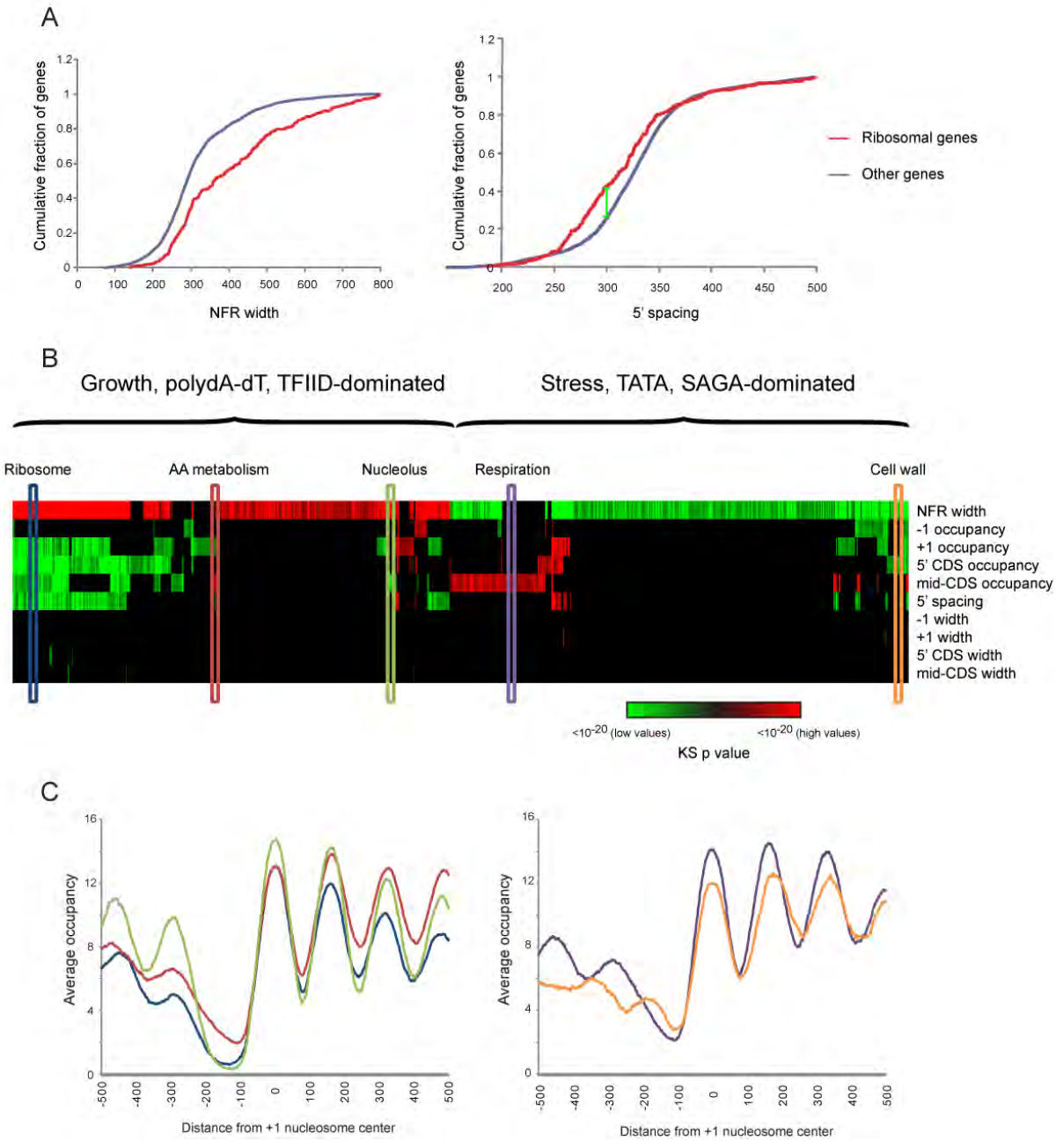
Genome-wide maps of nucleosomes in actively growing yeast have been the subject of a rage of recent studies (Albert et al. 2007; Field et al. 2008; Lee et al. 2007; Mavrich et al. 2008a; Shivaswamy et al. 2008; Yuan et al. 2005). To further evaluate our methodology, we carried out deep sequencing of mononucleosomal DNA from actively growing yeast, and applied our method to generate a map of nucleosome locations. Our method automatically extracts features of interest, such as NFR width, nucleosome length (which is not available using previous methods), nucleosome spacing, and occupancy. To systematically examine these features we used a compendium of experimental gene annotations we previously collected (Dion et al. 2007; Wapinski et al. 2007) and compared these against multiple nucleosome attributes associated with each gene (e.g., +1 nucleosome occupancy, spacing between the +1 and +3 nucleosome, etc.). Using the Kolmagorov-Smirnov test (Figure II.2A) we discovered attributes whose distribution in specific gene groups was significantly (FDR < 0.05) different from the background (Figure II.2B).

Figure II. 2: Different promoter types are differently-packaged

A) Cumulative distribution function (CDF) plots for two significant Kolmogorov-Smirnov enrichments. The geneset of 270 ribosomal genes is enriched for long NFRs (left panel), and close +1 to +3 nucleosome spacing (right panel). For example, 45% of Ribosomal genes have 5' nucleosome spacing of less than 300 bp (green line), whereas only 25% of all genes have this spacing.

B) Enrichment of high or low values of various parameters for a set of promoter types. Various parameters such as +1 nucleosome occupancy (listed at on right) were extracted for all yeast promoters. Each of the geneset previously gathered (Dion et al. 2007; Wapinski et al. 2007) was tested (using KS test) for significantly high or low values of the various parameters. Significant (FDR < 0.05) high values are shown in red, and significant low values are in green. Colors represent $\log_{10}$ of the p-value of the KS enrichment (saturated at p < 1e-20). Boxes indicate gene classes for Part C.

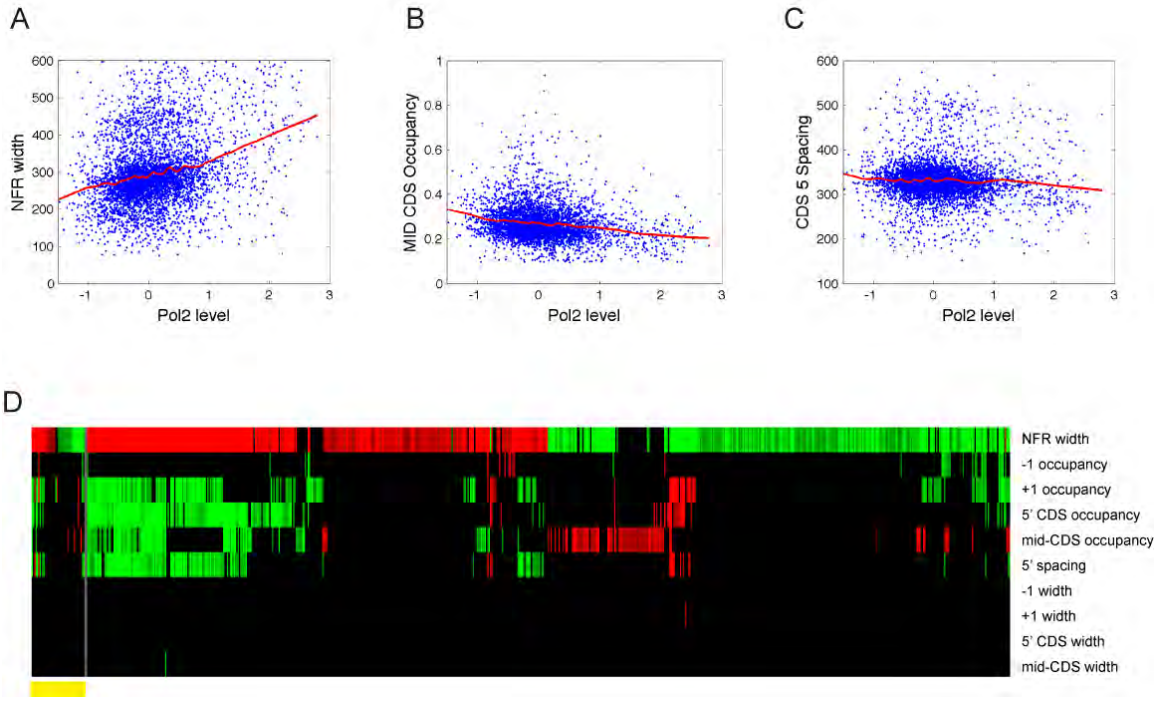
C) Averages for five promoter types as indicated in B, aligned according to +1 nucleosome center. Y axis represents average normalized nucleosome occupancy, in # reads per million mappable reads


A
Cumulative fraction of genes
NFR width
5' spacing
Ribosomal genes
Other genes
B
Growth, polydA-dT, TFIID-dominated
Stress, TATA, SAGA-dominated
Ribosome
AA metabolism
Nucleolus
Respiration
Cell wall
NFR width
-1 occupancy
+1 occupancy
5' CDS occupancy
mid-CDS occupancy
5' spacing
-1 width
+1 width
5' CDS width
mid-CDS width
<10-20 (low values)
<10-20 (high values)
KS p value
C
Average occupancy
Distance from +1 nucleosome center
Distance from +1 nucleosome center

This analysis highlights previously described features of yeast chromatin as well as novel ones. We recapitulate the dichotomy between promoter packaging of growth (TFIID-regulated, TATA-less, high expression growth genes) and stress (SAGA-regulated, TATA, noisy expression, rapid histone replacement) genes (Albert et al. 2007; Choi and Kim 2008; Choi and Kim 2009; Dion et al. 2007; Field et al. 2008; Ioshikhes et al. 2006; Newman et al. 2006; Tirosh and Barkai 2008) (Figure II.2B). For example, canonical growth genes, such as those encoding ribosome proteins, are enriched with wider NFRs and often also exhibit lower nucleosomal occupancy of coding regions, whereas stress genes, such as heat shock-induced genes, show the opposite behavior. This dichotomy is clearly the major feature that separates classes of genes in yeast, but our analysis also shows finer distinctions beyond this dichotomy. In general, the second distinguishing axis beyond NFR width is coding region nucleosome occupancy, which further can be separated into occupancy of 5' nucleosomes such as the +1 nucleosome, and occupancy of mid-CDS nucleosomes located distal from either end of the gene. For example, within the "growth" class of genes, ribosomal and metabolic genes can be roughly distinguished by the occupancy of coding region nucleosomes (Figure II.2B,C). While these distinctions appear subtle, it is important to note that they are of similar magnitude to physiologically-relevant (Ihmels et al. 2005) changes observed in the chromatin packaging of mitochondrial ribosomal genes between

S. cerevisiae and *C. albicans* (Field et al. 2009), suggesting that such small differences can indeed play roles in transcriptional control.

To explore the role of transcription in shaping chromatin architecture, we analyzed the relationship between Pol II enrichment (Methods) at a given gene and the various chromatin parameters described above (Supplementary Figure II.7). NFR width was positively correlated with transcription rate, while +1 occupancy, mid-CDS occupancy, and +1 to +3 spacing were slightly anticorrelated with transcription rate. To determine the extent to which gene set enrichments from Figure II.2A were driven by transcription level, we corrected each gene's chromatin parameters to account for transcription rate (Methods), and repeated the KS enrichment analysis from Figure II.2A (Supplementary Figure II.7D). The majority of enrichments from Figure II.2B repeated after correcting for Polymerase abundance (699 of 1001 retained, 302 lost, 42 gained), indicating that different regulatory mechanisms are linked to different chromatin architecture of gene sets.



Supplementary Figure II. 7: RNA polymerase accounts for some aspects of chromatin architecture

A-C) Gene-by-gene scatterplot of RNA polymerase occupancy (ChIP enrichment, x- axis) vs. NFR width (A), mid-CDS nucleosome occupancy (B), and +1 to +3 nucleosome spacing (C). LOWESS fit is shown in red.

D) Gene set enrichments for various chromatin features are largely maintained after correcting for Pol2 occupancy. Extracted chromatin parameters for all genes were corrected for influence of Pol2 levels, and KS enrichment analysis from Figure II.2B was repeated. ~70% of enrichments are maintained, and only a small number (yellow bar) of enrichments were gained.

Analysis of Nuclease Titration Levels

A striking aspect of our analysis is that the majority of nucleosomes were matched using templates with multiple ends, suggesting that the majority of nucleosome ends are partially trimmed during a typical MNase digestion. Furthermore, in comparing our data to published datasets we found global

variation in the distribution of nucleosome widths between datasets (Supplementary Figure II.3B), as well as in the relative occupancy of various nucleosome classes (ie genome-wide averages of +1 occupancy vs. mid-CDS occupancy differs between our data and that of Shivwaswamy et al (Shivaswamy et al. 2008), not shown). Indeed, a recent study using MNase titrations followed by q-PCR identified variation in quantitative MNase susceptibility across nucleosomes associated with *GAL* genes (Bryant et al. 2008). We therefore sought to more thoroughly explore the influence of digestion level in nucleosome positioning and occupancy. We have previously reported little change in nucleosome maps as measured by tiling microarray when mononucleosomal DNA is isolated from an early digestion step with only ~40% mononucleosomal DNA (Yuan et al. 2005). However, dynamic range compression by microarrays might hide changes in relative abundance of nucleosomes, and we were unable to obtain enough DNA for microarray analysis from less-digested (<40% mononucleosome) titration steps. Moreover, small changes in nucleosome segment lengths are difficult to detect with tiling microarrays.

Since limited digestion with trypsin has proven a valuable structural probe for proteins, we were also interested in whether the same might be true of limited nuclease digestion of chromatin. We therefore carried out a titration of micrococcal nuclease, and gel-purified mononucleosomal DNA from three different titration levels – underdigested (~15% mononucleosomes, “BY2”), typical digestion (~80% mononucleosomes, “BY10”), and overdigested (only

mononucleosomal DNA, “BY15”) (Figure II.3A). We used Template Filtering to call nucleosome positions in our titration data. As expected, nucleosome length was correlated with digestion level (Figure II.3B), with increasing digestion leading to shorter and shorter nucleosomes, presumably due to “chewing” of nucleosome ends by MNase.

Inspection of nucleosome maps for the three digestion levels revealed extensive similarities between the three maps. However, notable changes occur, particularly between under-digested and typical digestion (Figure II.3C). To globally assess differences between the different titration steps, we aligned genes by transcriptional start site (Figure II.3D) or stop codon (Figure II.3E) and averaged data from all genes at the 3 different titration levels. At 5’ ends of genes, we found an anti-correlation between +1 nucleosome occupancy and digestion level as expected. +1 nucleosomes are most abundant in underdigested chromatin, and least abundant in overdigested chromatin. This is seen both in TSS-aligned averages of all genes (Figure II.3D), as well as in systematic analysis of changes in relative nucleosome occupancy calls (Supplementary Figure II.8).

Figure II. 3: Effects of MNase level on chromatin structure

A) Mononucleosomal DNA was isolated from ladders from three different MNase titration levels, and sequenced by Illumina sequencing.

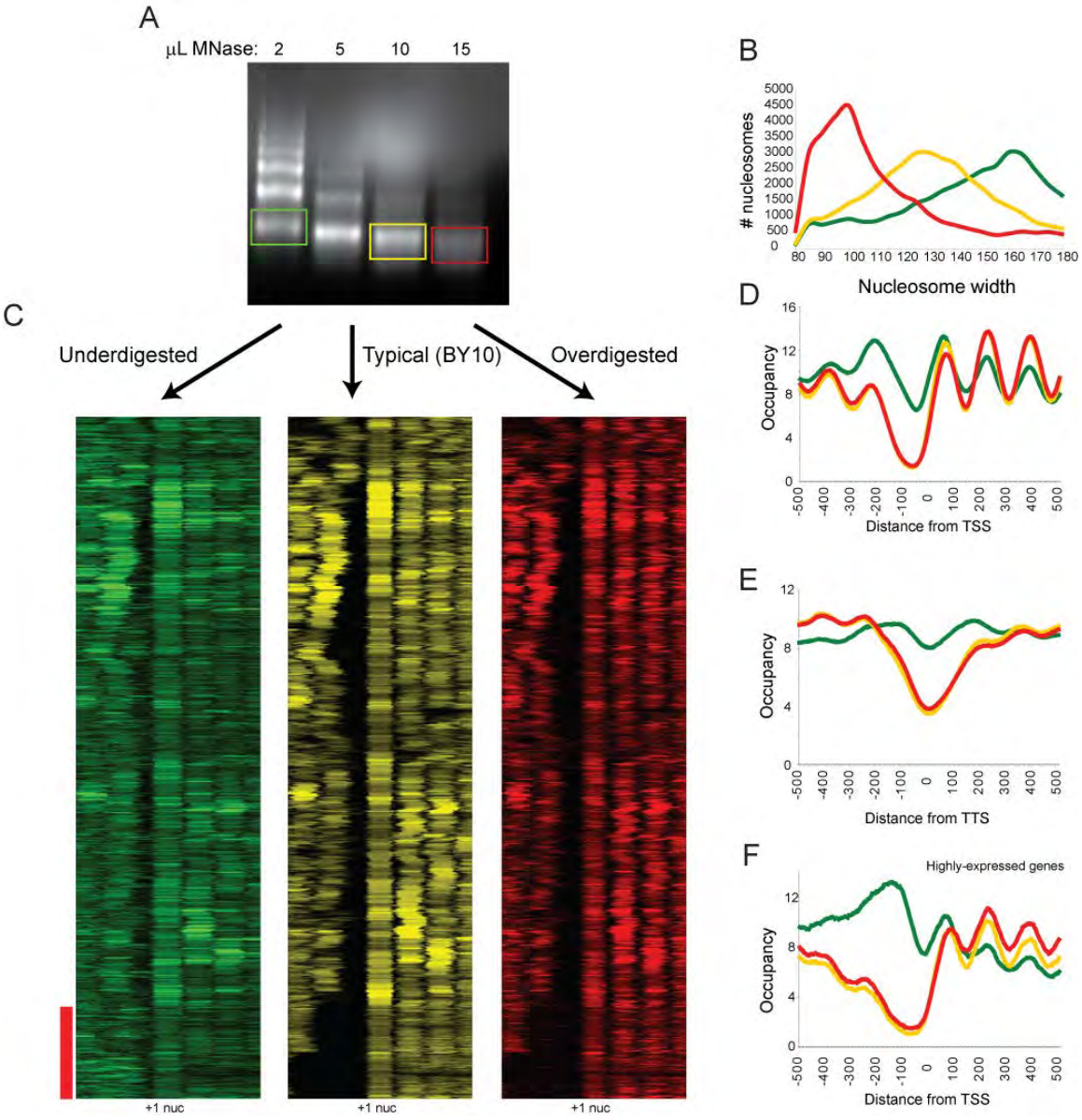
B) Data from titration series was subjected to Template Filtering to generate nucleosome calls. Width distributions for nucleosomes from the three titration steps are plotted. Green, yellow, and red correspond to under-, mid-, and over- digested chromatin, respectively.

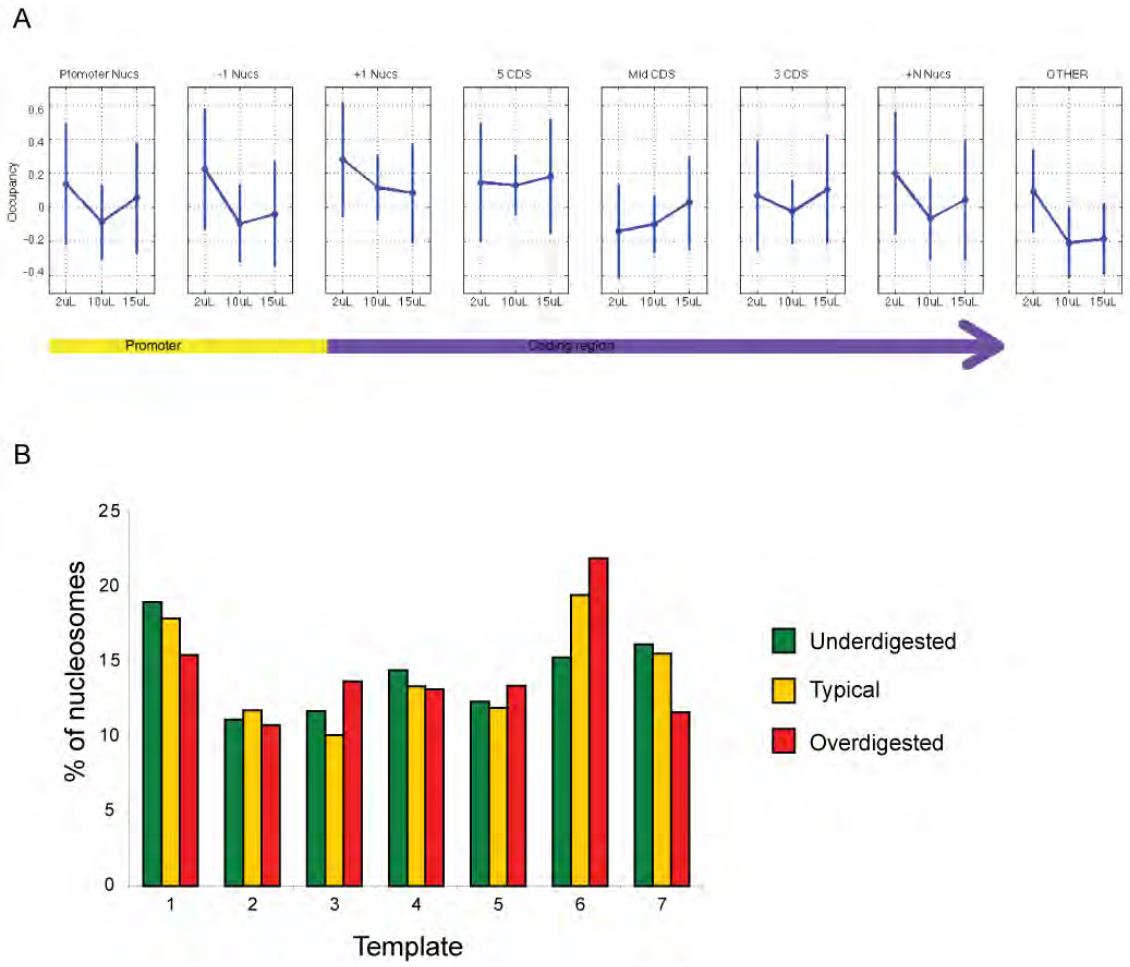
C) Data for under- (green), mid- (yellow), and over-(red) digested chromatin is shown in cluster view. Genes are aligned using BY10 +1 nucleosome center (indicated), all three clusters have genes ordered by clustering for BY10 data. Red bar indicates genes with wide NFRs in mid- and over-digested chromatin (largely highly-expressed genes such as ribosomal genes), that are partially filled in under-digested chromatin.

D) TSS-aligned nucleosome occupancy data for all genes.

E) Stop-codon-aligned nucleosome occupancy for all genes.

F) As in D, but only for genes with Pol2 ChIP occupancy > 1, top 7% of genes.


A
µL MNase: 2 5 10 15
B
nucleosomes
Nucleosome width
C
Underdigested
Typical (BY10)
Overdigested
+1 nuc
+1 nuc
+1 nuc
D
Occupancy
Distance from TSS
E
Occupancy
Distance from TTS
F
Highly-expressed genes
Occupancy
Distance from TSS



Supplementary Figure II. 8: Effects of MNase level on nucleosome occupancy and template usage

A) Occupancy of nucleosomes at varying locations relative to coding regions is shown for BY2, BY10, and BY15 titration levels.

B) Percentage of nucleosomes matching templates 1 to 7 is shown for BY2, BY10, and BY15.

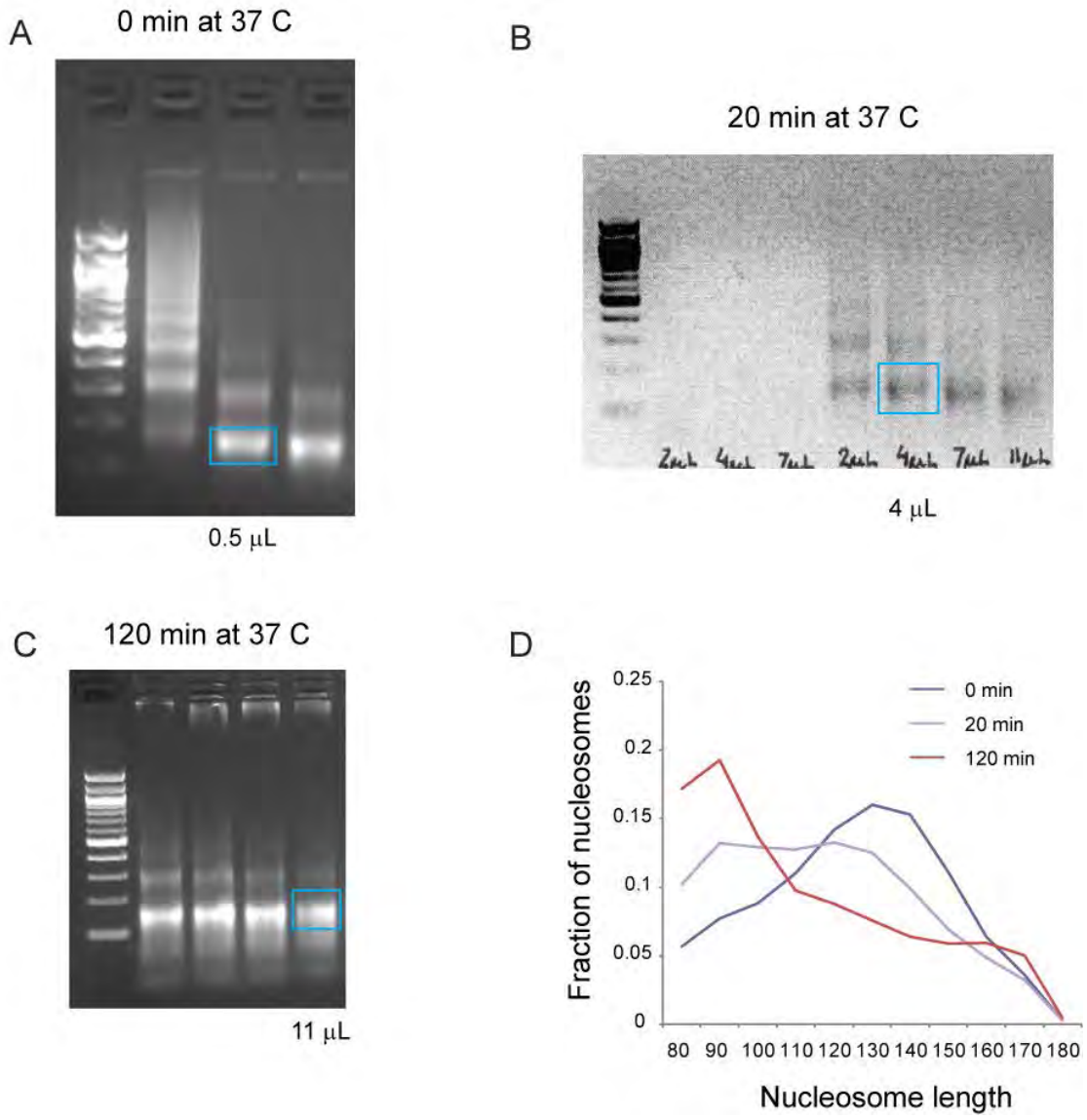
The loss of promoter-proximal nucleosomes during digestion was even more pronounced for -1 nucleosomes, where underdigested chromatin (BY2) showed high levels of the -1 (and a resulting decrease in the width of the average NFR), whereas this nucleosome was less abundant or completely missing in BY10 and BY15 (Figure II.3C, red bar). A similar effect was observed at the 3' NFR (Figure II.3E). The presence of an easily-digested nucleosome over the NFR was most commonly observed at the long NFRs associated with highly-expressed genes (red bar), and can easily be seen in averaged data for genes with the highest levels of RNA polymerase (Figure II.3F). Thus, consistent with recent studies in Drosophila (Henikoff et al. 2009) and human (Jin et al. 2009) cells, we find that at least some of the "nucleosome-free" region seen in typical nucleosome mapping studies corresponds to a loosely-bound (as determined by salt extraction in those studies), easily-digested (seen here) nucleosome. Overall, nucleosome occupancy levels are most even in BY10, which is the digestion level we typically use for nucleosome mapping (Yuan et al. 2005), and histone modification mapping (Liu et al. 2005).

Together, the results of the titration series suggest the presence of easily-digested nucleosomes or other protein complexes at the promoters of highly expressed genes, and point towards the necessity of knowing the extent of digestion when comparing nucleosome maps from different labs or different experiments.

The Role of RNA Polymerase in Chromatin Structure

A number of recent studies have claimed that intrinsic affinity of various genomic sequences for the histone octamer accounts for much, or most, of the chromatin structure observed in vivo in yeast (Field et al. 2008; Ioshikhes et al. 2006; Kaplan et al. 2008; Peckham et al. 2007; Segal et al. 2006; Yuan and Liu 2008). However, experimental determination of nucleosome positioning after in vitro reconstitution revealed that intrinsic preferences can almost entirely be ascribed to the role of polydA/dT in excluding nucleosomes, with little additional translational positioning information encoded in the genome (Kaplan et al. 2008; Zhang et al. 2009). The huge discrepancy between in vitro sequence preferences and in vivo nucleosome positioning is likely to result from the action of numerous factors, most notably protein complexes in vivo that move nucleosomes from their preferred positions, such as the ATP-dependent chromatin remodeler Isw2 (Whitehouse et al. 2007; Whitehouse and Tsukiyama 2006). Almost certainly the most widespread of these trans-acting factors is RNA polymerase II, as ~2/3 of the yeast genome codes for proteins, and combining distributions of RNA abundance (Yassour et al. 2009) with typical absolute mRNA abundances (Iyer and Struhl 1996) and half-lives (Wang et al. 2002) indicates that well over half the yeast genome is likely transcribed at least once during a given cell cycle. The passage of RNA polymerase disrupts DNA-histone contacts (Wasylyk and Chambon 1980), leading us to ask how RNA polymerase globally affects chromatin structure in vivo.

We examined the effects of RNA polymerase on chromatin structure by using the *rpb1-1* yeast strain, which contains a temperature-sensitive allele of the gene encoding the large subunit of RNA Polymerase II (Nonet et al. 1987). To identify both the early as well as longer term effects of RNA polymerase deactivation, we performed MNase-seq at 0, 20, and 120 minutes after shifting these cells from 25 C to 37 C. Interestingly, we found that despite minimal change in cell density, increasing amounts of MNase were required to generate similar nucleosome ladders (Supplementary Figure II.9A-C) as the time course progressed, indicating a global role for RNA polymerase in increasing overall chromatin accessibility. Consistent with the increased MNase required, we found that nucleosome length decreased over the time course (Supplementary Figure II.9D). Importantly, none of our conclusions below are affected by the titration level, as the inferred effects of Polymerase loss do not match effects of overdigestion (see below).



Supplementary Figure II. 9: Polymerase inactivation results in globally inaccessible chromatin

A-C) Gels from MNase titrations for *rpb1-1* cells grown at 25 C (A), or shifted to 37 C for 20 minutes (B), or 120 minutes (C). Note the increasing amounts of MNase required to achieve the same nucleosome ladder.

D) Nucleosome length distributions for the three nucleosomal populations sequenced (boxes in parts A-C). Consistent with the increased MNase used at later time points, nucleosome widths decrease.

Averaging genes by TSS for the three time points reveals two effects of polymerase loss on nucleosome positioning (Figure II.4A,B, S10). First, NFR width decreases over time, largely because of gains in nucleosome occupancy at the -1 position (Venters and Pugh 2009). This was not a digestion artifact at later time points, as the nucleosome length distribution at 120 minutes was consistent with MNase *over*digestion (Supplementary Figure II.9D), whereas increased -1 nucleosome occupancy is a general property of *under*digested chromatin (Figure II.2). Second, coding region nucleosomes shift downstream over time. This latter observation is interesting, as it is consistent with predictions from biochemical studies: the process of RNA polymerase transiting a nucleosome in vitro results in transient dissociation of 5' DNA from the octamer, followed by recapture of upstream DNA on the octamer surface, resulting in a predicted retrograde shift of the octamer relative to Polymerase movement (Studitsky et al. 1994; Studitsky et al. 1997).

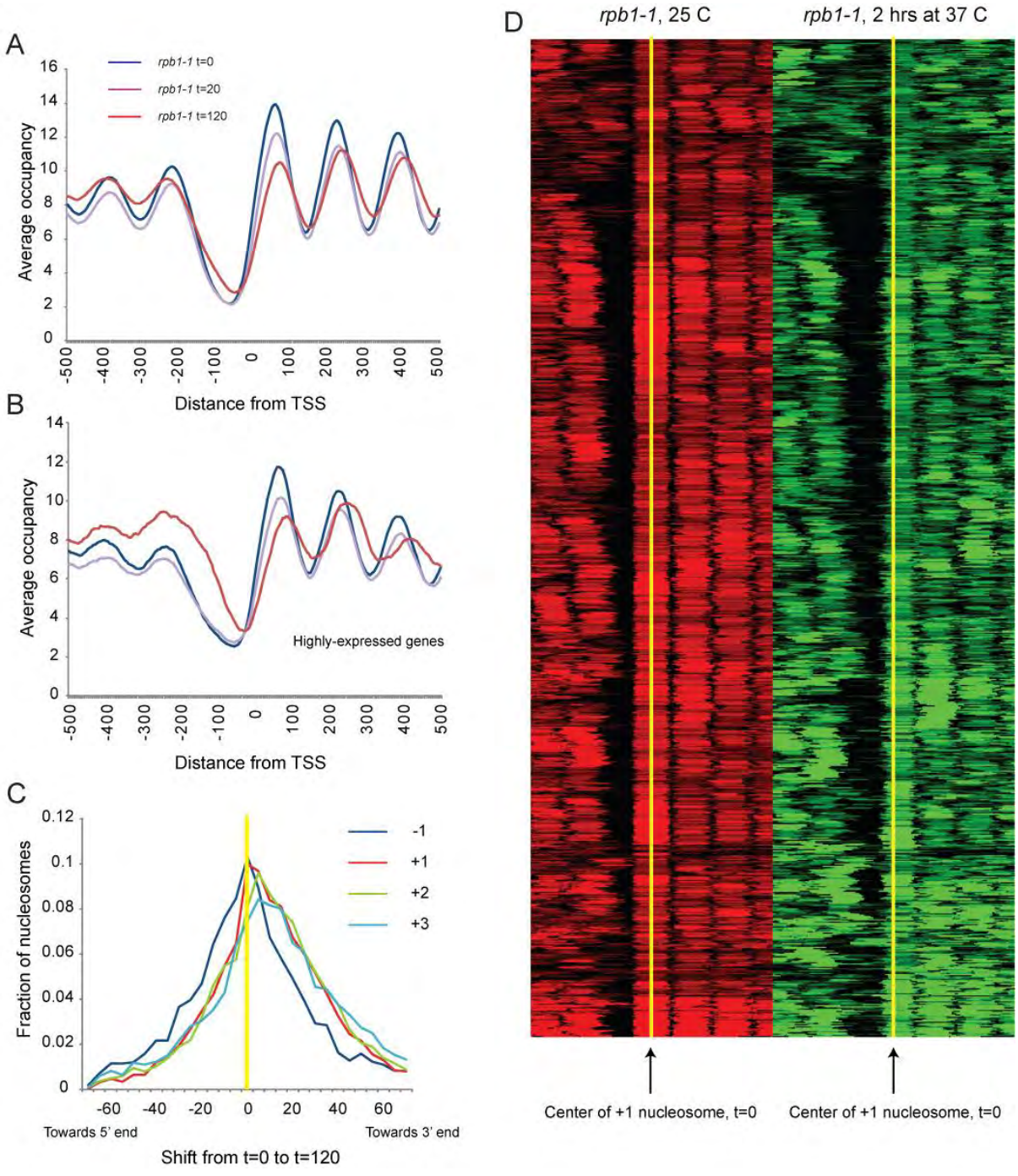
Figure II. 4: Effects of RNA polymerase on chromatin structure

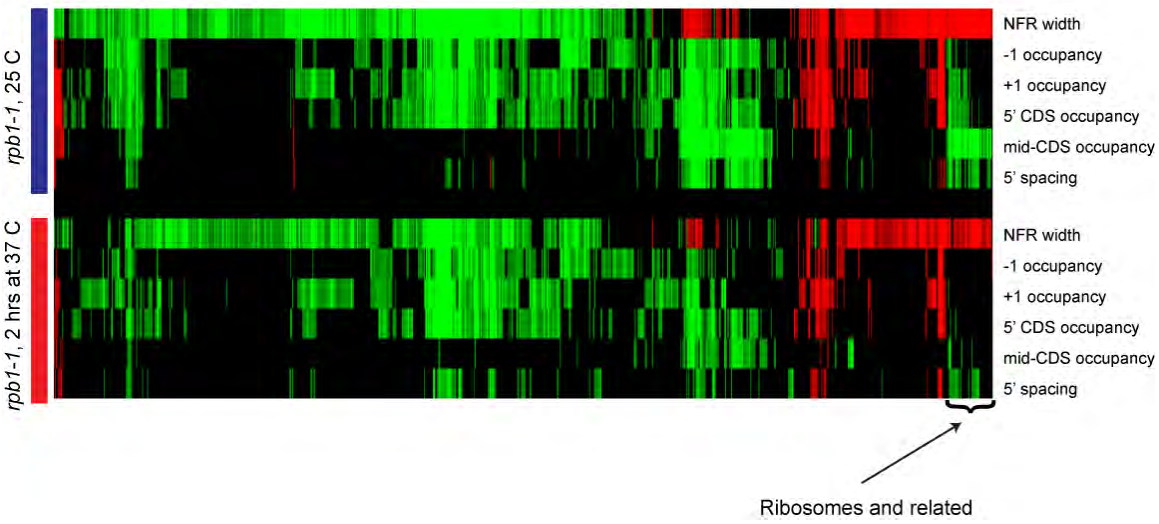
A) Nucleosomes were isolated from *rpb1-1* yeast grown at 25 C, and shifted to 37C for 20 or 120 minutes. Data are presented in TSS-aligned average.

B) As in A), but for highly-expressed genes.

C) Nucleosomes over genes shift downstream upon Pol2 loss. For each indicated nucleosome type (-1, +1, +2, +3) we plot the distribution of center-to-center distances between the nucleosome calls at 0 and 120 minutes after Pol2 inactivation. We find that 43% of -1 nucleosomes, 59% of +1 nucleosomes, 61% of +2 nucleosomes, and 60% of +3 nucleosomes shift away from the NFR.

D) Global view of +1 nucleosome shifts during Pol2 inactivation. Nucleosome calls for all promoters with a downstream +1 nucleosome shift are shown as a heatmap, aligned by the center of the +1 nucleosome (yellow) before Pol2 inactivation (red). After 2 hours of Pol2 inactivation, downstream shifts of these 59% of +1 nucleosomes are apparent.


A
rpb1-1 t=0
rpb1-1 t=20
rpb1-1 t=120
Average occupancy
Distance from TSS
B
Average occupancy
Highly-expressed genes
Distance from TSS
C
Fraction of nucleosomes
-1
+1
+2
+3
Towards 5' end
Towards 3' end
Shift from t=0 to t=120
D
rpb1-1, 25 C
rpb1-1, 2 hrs at 37 C
Center of +1 nucleosome, t=0
Center of +1 nucleosome, t=0

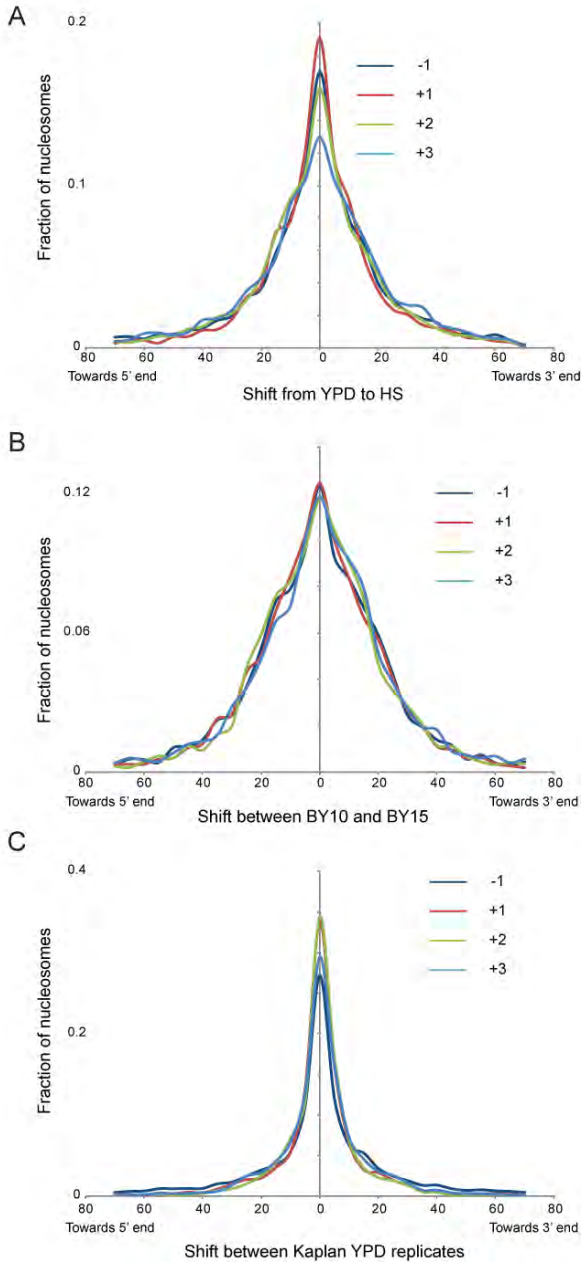


Supplementary Figure II. 10: Effects of RNA polymerase inactivation on gene category chromatin packing

Gene sets were assessed for enrichment of different chromatin parameters as in Figure II.2B, for *rpb1-1* yeast grown at 25 C or 37 C for 2 hours. Genesets are ordered as in Figure II.2B.

We investigated nucleosome position shifts by plotting the direction of nucleosome shift between pairs of time points during the temperature shift (Figure II.4C,D). Nucleosome shifts over coding regions were clearly biased in the downstream direction, consistent with the averaged view in Figure II.4A. Importantly, -1 nucleosome shifts were not biased in either direction. This small (~10 bp on average) downstream shift is specific, as comparisons between other pairs of chromatin datasets (such as pre- and post-heat shock) show distributions of nucleosome shifts centered on zero (Supplementary Figure II.11). To

determine whether nucleosome shifts preferentially occurred over particular classes of genes we tested gene classes for significant deviation from the average shift of the -1, +1, +2, or +3 nucleosome by the KS statistic. Again, very highly-expressed gene classes such as those encoding ribosomal proteins or amino acid metabolism genes exhibited more dramatic nucleosome shifts from 20 to 120 minutes (not shown). Highly-expressed genes are generally downregulated during heat stress (Gasch et al. 2000), but nucleosome shifts at ribosomal genes are unlikely to be a consequence of heat shock-induced changes – re-analysis of the heat shock data from Shivaswamy et al (Shivaswamy et al. 2008) do not recapitulate the shifts we observe here (Supplementary Figure II.11A).



Supplementary Figure II. 11: Downstream nucleosome shifting is specific to Pol2 shutoff

A) Analysis as in Figure II.5C, comparing nucleosome calls from pre- and post-heat shock conditions from Shivaswamy et al. Note that coding region nucleosomes show no bias for downstream shifts during heat stress.

B) As above, for BY10 and BY15 data from this study.

C) As above, for replicate in vivo data (YPD) from Kaplan et al.

Why do nucleosomes shift after loss of polymerase? As noted above, RNA polymerase is a major factor in nucleosome movement and eviction in vivo, and the in vivo nucleosome positions over highly-expressed genes deviate from in vitro nucleosome preferences to a greater extent than they do over poorly-expressed genes. Thus, we compared our data to the in vitro data from Kaplan et al (Kaplan et al. 2008), reasoning that loss of RNA polymerase might allow chromatin to relax to more closely match local thermodynamic minima. We noticed that nucleosomes at many promoters indeed more closely match in vitro preferences after 2 hours of polymerase inactivation (Figure II.5A).

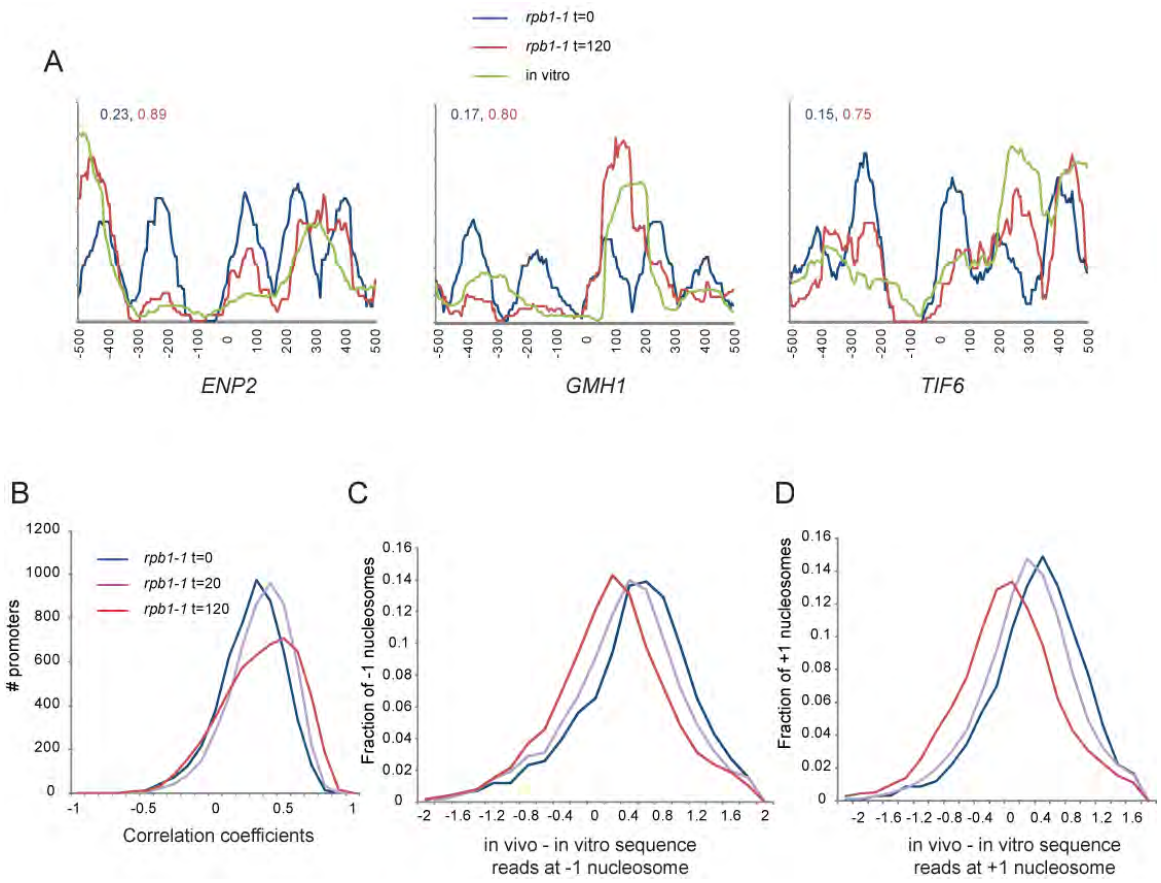


Figure II.5: Nucleosomes relax towards in vitro preferred locations after Pol2 loss

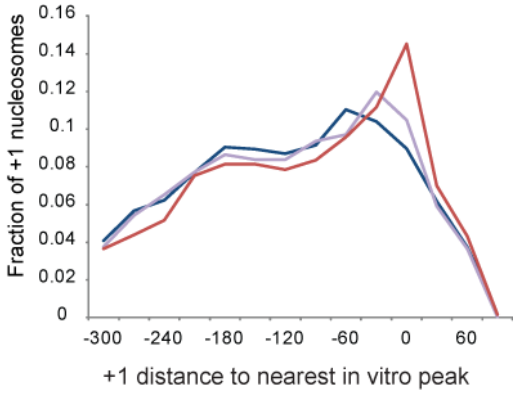
A) Three examples of promoters where data from Pol2 inactivation matches in vitro nucleosome assembly data better than data from before Pol2 inactivation. Shown are extended read coverage along 1000 bp centered on TSS. Numbers shown in inset are correlations between in vitro coverage and t=0 (blue) and t=120 (red) in vivo coverage.

B) Promoter chromatin architecture globally shifts towards in vitro preferences as polymerase is inactivated. Extended read coverage along the 1 kb centered on the TSS was extracted for all genes, and correlation coefficients were calculated to equivalent data for in vitro nucleosome reconstitution experiments (Kaplan et al. 2008). Histograms show a global shift of promoters towards the in vitro nucleosome pattern.

C) Normalized occupancy of -1 nucleosome better matches in vitro data after polymerase loss. For all -1 nucleosomes (called at t=0 or at t=120), the difference between in vivo normalized occupancy and in vitro normalized occupancy at the center of the in-vivo nucleosome were calculated and presented as a histogram.

D) As in C, but for all +1 nucleosomes.

To generalize this result, we calculated the correlation between in vitro nucleosome assembly data (Kaplan et al. 2008) and in vivo nucleosome positioning for 1 kb windows at the 5' ends of genes (Figure II.5B). At t=0, correlation coefficients centered around 0.3, consistent with the ability of in vitro assembly to highlight NFRs (Kaplan et al. 2008; Sekinger et al. 2005; Zhang et al. 2009). After 2 hours of polymerase inactivation, the distribution of correlations shifted to a higher value of ~0.5, indicating that RNA polymerase does help maintain nucleosomes in thermodynamically-unfavored locations in vivo. Of course, much of this is due to the above-mentioned role of transcription in -1 nucleosome eviction (Figure II.5C). We also asked whether lateral repositioning of nucleosomes upon polymerase loss resulted in relaxation to thermodynamically-preferred positions. Strikingly, distances between +1 nucleosomes and the nearest *in vitro* occupancy peak decreased after polymerase inactivation (Figure II.5D, Supplementary Figure II.12), consistent with the hypothesis that both nucleosome eviction and sliding by RNA polymerase antagonize the thermodynamically-preferred chromatin state.



Supplementary Figure II. 12: Polymerase loss results in nucleosome shifts towards in vitro peaks

For each +1 nucleosome, the distance from the nucleosome center to the nearest local maximum from Kaplan et al's in vitro reconstitution data (Kaplan et al. 2008) was calculated, and distances are presented in histograms for different times after polymerase inactivation.

DISCUSSION

Genome-wide mapping of nucleosome positions in *S. cerevisiae* has been a tremendously productive method for illuminating the principles underlying chromatin structure and function. Deep sequencing methods have multiple advantages over tiling microarrays for genomic localization studies such as nucleosome mapping studies, including single nucleotide resolution, expanded dynamic range, and nearly whole-genome coverage. Here, we present a novel method for analyzing deep sequencing data for chromatin maps. Our method automatically extracts nucleosome position, occupancy, and width, and accounts for variability in end digestion by MNase.

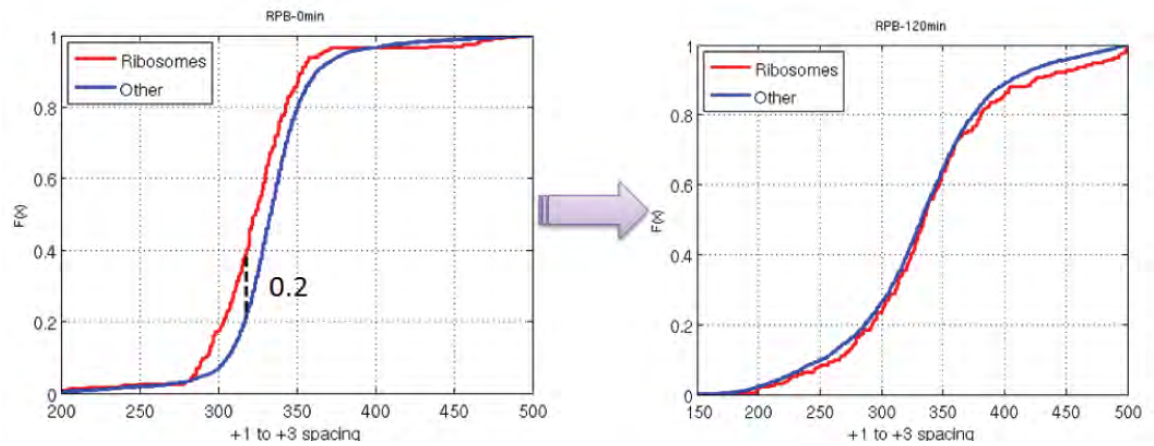
Analysis of chromatin packaging across the yeast genome confirmed previously-described aspects of yeast chromatin structure, including widespread 5' and 3' nucleosome-depleted regions, a dichotomy between stress and growth genes reflected in NFR width, increased nucleosome fuzziness distal to the NFR, and a subtle anticorrelation between coding region nucleosome occupancy and transcription. We also identified finer distinctions between certain classes within the major stress/growth branches. Most interestingly, we found that +1 to +3 nucleosome spacing was significantly shorter over ribosomal genes than over other gene types (see below).

We also analyzed data from an MNase titration series, as different laboratories isolate nucleosomes from different MNase digestion levels (see, for example, Shivwaswamy et al (Shivaswamy et al. 2008)). Analysis of data from

underdigested chromatin revealed the abundant presence of nucleosome-sized peaks in the “NFR”. A recent analysis of Drosophila chromatin also identified nucleosomes in NFRs in a low-salt extraction from underdigested chromatin (Henikoff et al. 2009), and similar results hold in human cells (Jin et al. 2009). Does this material correspond to easily-digested nucleosomes, or to DNA protected from MNase by other proteins such as transcription factors? Two lines of evidence support the former hypothesis. One, the equivalent material in Drosophila is associated with the histone variant H3.3, indicating the presence of histones at these locations (Henikoff et al. 2009). Second, we only find nucleosomes filling in NFRs that are larger than 140 bp, suggesting that these are bona fide nucleosomes. Interestingly, analysis of sequence motifs in that exhibit occupancy differences include the CGCG motif recently shown to be bound by the Rsc3/30 subunits of the RSC chromatin remodeling complex (Badis et al. 2008), suggesting that the easily-digested -1 peak might correspond to a nuclease-accessible RSC-remodeled nucleosome state.

As many features of yeast chromatin correlate with transcription rate, we mapped nucleosomes before and after inactivation of Pol II. We find that NFRs become shorter and shallower upon loss of Pol II, particularly at highly-expressed genes, consistent with a previously-described role for RNA polymerase in eviction of -1 nucleosomes (Venters and Pugh 2009). We also found a surprising, general role for RNA polymerase in nucleosome sliding – nucleosomes over coding regions generally shifted away from the NFR upon loss of polymerase

(Figure II.4). These results are consistent with the predictions from biochemical studies – in vitro, RNA polymerase is capable of transiting a nucleosome without evicting histones, apparently by invading a nucleosome edge and then propagating a bubble of DNA around the octamer surface (Hodges et al. 2009; Studitsky et al. 1994; Studitsky et al. 1997). This retrograde nucleosome movement may play a role in the stereotyped +1 nucleosome positioning in vivo, which is not explained by intrinsic thermodynamic preferences as measured by in vitro nucleosome assembly (Kaplan et al. 2008; Zhang et al. 2009). We speculate that after assembly of a newly-replicated DNA into nucleosomes, polymerase passage could be responsible for nucleosome shifts towards the NFR until the +1 nucleosome is as close to either polyAs or to the preinitiation complex as physically possible. This retrograde nucleosome movement may also play a role in the surprisingly tight packing of nucleosomes over highly-transcribed ribosomal genes, and indeed loss of polymerase results of relaxation of the +1 to +3 spacing at these genes (Supplementary Figures II.10, II.13).



Supplementary Figure II. 13: Polymerase loss eliminates short +1 to +3 spacing at ribosomal genes

Plots as in Figure II.2A for +1 to +3 nucleosome spacing for *rpb1-1* yeast grown at 25 C (left) or at 37 C for 2 hours (right).

This interpretation must be tempered by dynamic studies in yeast, which indicate that some nucleosomes (particularly +1 nucleosomes) are rapidly exchanged during G1 (Dion et al. 2007; Jamai et al. 2007; Rufiange et al. 2007) – how is it that translational effects of transcription are observed on nucleosomes given that nucleosomes are often rapidly-exchanged (in some cases many times per cell cycle)? We consider two of many possible ways to reconcile these results. First, given that current locus-specific dynamic exchange measurements rely on transcriptional activation of tagged histones, there is a lower bound (~15-30 minutes) for the fastest exchange rates measurable. However, if the fastest exchange rates are indeed on the order of ~15 minutes, then for many genes a

high proportion of cells in a population will have had polymerase pass through the gene since the last histone turnover cycle, resulting in a detectable polymerase-driven retrograde shift in the population measurement despite ongoing nucleosome replacement. Second, we do not currently know the extent of correlation in the dynamics of adjacent nucleosomes at the single-gene level. In other words, at a highly-transcribed gene with high levels of histone replacement throughout the gene body, does +1 eviction affect +2 eviction? If these do not always co-occur, then retrograde shifts of surrounding nucleosomes could provide a local “memory” of prior Polymerase passage, such that a leftward-shifted +2 would constrain the replacement location for a replaced +1 nucleosome.

Finally, our results bear on the relationship between thermodynamic sequence preferences and in vivo chromatins structure. Dramatic claims have been made regarding the extent to which genomic sequence dictates the positioning of nucleosomes in the cell (Kaplan et al. 2008; Segal et al. 2006), although most studies find the effects of sequence on chromatin architecture to be modest (Ioshikhes et al. 2006; Peckham et al. 2007; Yuan and Liu 2008). While in vitro chromatin assembly correlates well with in vivo nucleosome positions (Kaplan et al. 2008; Zhang et al. 2009), this almost entirely results from the depletion of nucleosomes over polyA and related sequences (Drew and Travers 1985; Iyer and Struhl 1995; Kunkel and Martinson 1981; Sekinger et al.

2005). Our results confirm the expected role for RNA polymerase in movement of nucleosomes away from thermodynamically-preferred positions.

Together, these results further emphasize the role for RNA polymerase in shaping the chromatin landscape of the genome, and point towards the difficulty in disentangling cause and effect in the relationship between chromatin and transcription.

Materials and methods

Nucleosome isolation

Yeast culture, fixation, and MNase titrations were carried out as previously described (Yuan et al. 2005). For *rpb1-1* temperature shifts, cells were grown to an OD of 0.6 in YPD at 25 C, then culture aliquots were immediately shifted to 37 C by addition of an equal volume of YPD at 49 C. After recovery of digested DNA, mononucleosomal was gel-purified and subjected to Illumina sequencing as described in Shivwaswamy et al (Shivaswamy et al. 2008).

Data Availability

Sequencing data have been deposited at GEO, accession #GSE18530, and are available at our supplemental website as well: *http://compbio.cs.huji.ac.il/NucPosition*.

Template Filtering Algorithm

Using a sliding window across the genome, we cross-correlate each position with a pair of templates, one matching the forward reads and one the reverse reads. We enumerate all 7x7 possible combinations of Forward and Reverse templates. We repeat this scan with different spacing between both templates, to capture over and under digestion of the ~146bp nucleosomal DNA fragments. As a result, we obtain a correlation 'heat map' for each pair of templates containing the correlation coefficient for each center position and width. Next, we search for local maxima points within this 'heat maps'; each maxima point is a potential nucleosome at a given position with a specific width. Finally, to assemble the final set of nucleosmes, we are using a greedy approach to select the best assignment of nucleosomes under overlapping constrains. Potential nucleosomes are sorted according to correlation score and occupancy, and are then selected to the final set allowing maximum overlap of 40% between adjacent nucleosomes.

Selecting representative templates

To generate variety of templates that represents the prototypical distributions of reads at nucleosome ends, we first applied our method using a Gaussian shaped template and obtained a preliminary map of nucleosome predictions. We aligned all predicted nucleosome ends and created a matrix of read patterns using a window of 80 bp flanking the edges. Next, we clustered this matrix using k-mean clustering and selected seven representative templates that capture

Correcting chromatin parameters to account for Pol2 enrichment in K-S tests

We represent each gene as a vector of chromatin properties (i.e. +1 nucleosome occupancy, NFR width, mid CDS occupancy etc). Using a compendium of experimental gene annotations we previously collected (Dion et al. 2007; Wapinski et al. 2007) we compared the distribution of each chromatin parameter for each gene set vs. the background. We discovered 1001 gene sets with at least one enriched chromatin property. To correct the genes' properties vectors for RNA polymerase levels, we plotted RNA polymerase occupancy (measured by microarray as in Ref (Steinmetz et al. 2006), Kim et al., 2010) vs. each chromatin property and calculated the LOWESS curve for each property (Supplementary Figure II.7A-C). We then subtracted the smoothed LOWESS curves from each gene, obtaining new chromatin properties vectors that represent the distance of each property from the LOWESS curve. We repeated the K-S enrichment with these new vectors.

CHAPTER III: A functional evolutionary approach to identify determinants of nucleosome positioning: A unifying model for establishing the genome-wide pattern

ABSTRACT

Although the genomic pattern of nucleosome positioning is broadly conserved, quantitative aspects vary over evolutionary timescales. We identify the *cis* and *trans* determinants of nucleosome positioning using a functional evolutionary approach involving *S. cerevisiae* strains containing large genomic regions from other yeast species. In a foreign species, nucleosome depletion at promoters is maintained over poly(dA:dT) tracts, whereas internucleosome spacing and all other aspects of nucleosome positioning tested are not. Interestingly, the locations of the +1 nucleosome and RNA start sites shift in concert. Strikingly, in a foreign species, nucleosome-depleted regions occur fortuitously in coding regions, and they often act as promoters that are associated with a positioned nucleosome array linked to the length of the transcription unit. We suggest a three-step model, in which nucleosome remodelers, general transcription factors, and the transcriptional elongation machinery are primarily involved in generating the nucleosome positioning pattern *in vivo*.

INTRODUCTION

In living cells, nucleosome positions are influenced by intrinsic DNA sequence preferences due to the thermodynamic costs associated with wrapping stiff DNA around the histone octamer (Drew and Travers, 1985; Jiang and Pugh, 2009; Radman-Livaja and Rando, 2010). In addition, a wide variety of proteins can affect nucleosome positions and occupancy, most notably ATP-dependent chromatin remodeling complexes. The relative importance of DNA sequence and protein factors in determining nucleosome positioning has been subject to considerable debate. *In vitro* reconstitution of genomic DNA into nucleosomes by salt dialysis recapitulates gross variation in nucleosome occupancy in yeast and in humans – AT-rich sequences such as those found at yeast promoters are intrinsically nucleosome-depleted (Kaplan et al., 2009; Sekinger et al., 2005; Zhang et al., 2009), whereas the GC-rich sequences prevalent at human promoters are intrinsically nucleosome-enriched (Valouev et al., 2011). These studies typically find little role for intrinsic preferences in precise nucleosome positioning, although the enrichment of particular sequence features (10 bp periodicity of AA/AT/TA dinucleotides) at the +1 position in budding yeast has nonetheless led to forceful (Kaplan et al., 2010; Kaplan et al., 2009; Segal et al., 2006), but disputed (Fan et al., 2010; Stein et al., 2009; Weiner et al., 2010; Zhang et al., 2009, 2010) claims that intrinsic DNA sequence preferences play a major determining role in nucleosome positioning.

Conversely, several experimental approaches, largely in budding yeast, have revealed a key role for proteins in establishing nucleosome positions *in vivo*. While *in vitro* reconstitution of DNA into nucleosomes does not properly establish nucleosome positions at *PHO5*, addition of yeast whole cell extract enables more accurate assembly of nucleosomes at this locus (Korber and Horz, 2004). Genome-wide analysis subsequently showed that one or more ATP-dependent activities in yeast whole cell extract can assemble nucleosomes in positions that resemble, but do not completely coincide with, *in vivo* positioning (Zhang et al., 2011), thereby demonstrating a major role for nucleosome-remodeling complexes in nucleosome positioning. Decades of biochemical studies have identified many specific proteins and protein complexes capable of altering nucleosome positions on DNA *in vitro* (Clapier and Cairns, 2009), and increasingly these factors are being implicated in proper nucleosome positioning *in vivo* (Gkikopoulos et al., 2011; Whitehouse et al., 2007; Whitehouse and Tsukiyama, 2006). For example, the ATP-dependent remodeling enzymes Chd1, Isw1, and Isw2 globally affect nucleosome positioning *in vivo*, as their deletion in yeast leads to nearly complete loss of nucleosome positioning downstream of the +2 nucleosome of coding regions (Gkikopoulos et al., 2011).

In a genetic approach to this problem, diploid hybrids between the closely-related species, *S. cerevisiae* and *S. paradoxus*, have been used to determine to what extent divergent nucleosome positioning on specific orthologous genes can be attributed to *cis* or *trans* factors, with the majority of chromatin changes

between these species being attributed to poly(dA:dT) elements at promoters (Tirosh et al., 2010). However, *S. cerevisiae* and *S. paradoxus* differ very little in bulk aspects of chromatin architecture. In contrast, chromatin structure exhibits far greater differences between more divergent species: for example, average nucleosome spacing differs by ~15-20 bp between *S. cerevisiae* and *K. lactis* (last common ancestor ~150 MYA) (Heus et al., 1993; Tsankov et al., 2010).

Here, we describe a functional evolutionary approach to systematically dissect the contributions of DNA sequence and the nuclear environment to nucleosome positioning *in vivo*. This approach relies on the finding that there are species-specific differences in parameters of nucleosome positioning in a variety of yeast species, even though the general pattern is highly conserved (Tsankov et al., 2010). Specifically, we compare nucleosome maps of artificial chromosomes (YACs) containing large, heterologous genomic regions from different yeast species in *S. cerevisiae* with maps of the same regions in their native organism (Figure III.1A). In principle, features that change in the context of *S. cerevisiae* are determined by protein factors that are functionally distinct in the two species, whereas features that are retained when the foreign yeast DNA is present in *S. cerevisiae* are either due to intrinsic DNA sequence or to conserved *trans*-acting regulators. For example, when the *S. cerevisiae HIS3-PET56* region is introduced into *S. pombe*, it retains the nucleosome-depleted promoter region, but not the positions of nucleosomes in the coding region (Sekinger et al., 2005). In addition, the generation of fortuitous functional elements arising from

heterologous genomic sequences makes it possible to address mechanistic issues that are presumably free of evolutionary constraints. Here, we show that nucleosome spacing is established in *trans*, and that promoter nucleosome depletion can be established either by intrinsic sequence cues or by *trans*-acting factors. Further, we find that +1 nucleosome positioning is most likely established by some aspect of the transcriptional machinery, and positioning of more downstream nucleosomes in the mRNA coding region is linked to Pol II elongation. Based on results presented here and elsewhere, we propose a unifying, three-step model for how nucleosome positions are established *in vivo*.

RESULTS

Generation of *S. cerevisiae* strains harboring artificial chromosomes with large segments of foreign yeast DNA

To generate yeast artificial chromosomes (YACs, Figure III.1A), genomic DNA from *K. lactis*, *K. waltii*, and *D. hansenii* was sheared to ~100-200 kb average size, and ligated to the pYAC4 vector carrying sequences for *S. cerevisiae* telomeres, centromere, and origin of replication, as well as two selectable markers. YACs were transformed into wild-type *S. cerevisiae* and confirmed by pulsed-field gel electrophoresis (Figure III.1B). Furthermore, both ends of YACs containing foreign yeast DNA inserts were validated by DNA sequencing. In total, we generated seven strains carrying distinct YACs from the three species, with an average insert length of ~140 kb (Table III.1). YAC strains were grown in identical conditions to those previously used for mapping nucleosomes in these

four species (Tsankov et al., 2010), and formaldehyde cross-linked chromatin was digested to ~80% mononucleosomes using micrococcal nuclease (Yuan et al., 2005). Mononucleosomal DNA was analyzed by deep sequencing as previously described (Shivaswamy et al., 2008; Tsankov et al., 2010; Weiner et al., 2010). Figures III.1C-D show nucleosome mapping data for two genes from *K. lactis*, with data from wild-type *K. lactis* in blue ("endogenous"), and data for these same genes in the context of a YAC-carrying *S. cerevisiae* strain in red ("YAC"). Notable in these views are a number of well-described aspects of fungal chromatin structure – in the endogenous context, nucleosomes are generally well-positioned (nucleosome peaks are well separated and exhibit high peak to trough ratios), and both genes have a nucleosome-depleted region (NDR) that contains the gene's promoter.

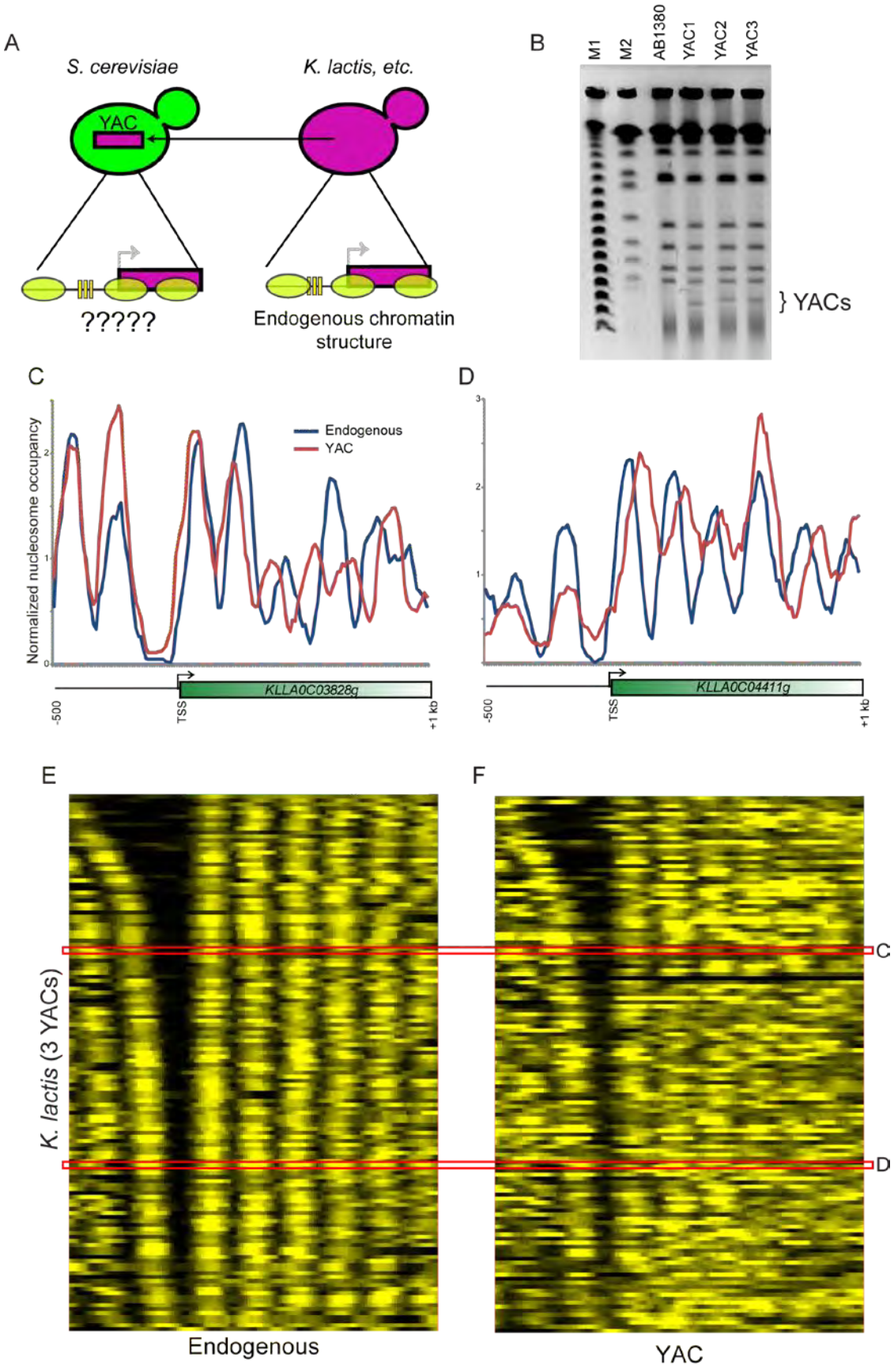
Figure III. 1: Functional evolutionary dissection of chromatin establishment mechanisms

(**A**) Schematic of experimental design. Yeast Artificial Chromosomes are constructed carrying sequence from species such as *K. lactis*, and introduced into *S. cerevisiae*. Comparison of nucleosome mapping data between the same sequence in two different environments (its endogenous genome, and in *S. cerevisiae*) can be used to disentangle DNA-driven from *trans*-mediated aspects of chromatin organization.

(**B**) Chromosomal complement of parental *S. cerevisiae* (AB1380) and 3 different YAC-bearing strains. Pulsed field gel electrophoresis of YAC-bearing strains, as indicated.

(**C-D**) Examples of nucleosome mapping data from two genes. Blue line indicates nucleosome mapping data from wild-type *K. lactis* (Tsankov et al., 2010), red line shows data from the same sequence carried on a YAC in *S. cerevisiae*.

(**E-F**) Data for all *K. lactis* genes on all 3 YACs. (**E**) shows data for all genes from wild-type *K. lactis*, with genes sorted by NDR width, while (**F**) shows data from these genes on YACs, sorted identically. Black indicates no sequencing reads, yellow intensity indicates number of sequencing reads. C and D indicate the example genes shown above.


A
S. cerevisiae
K. lactis, etc.
YAC
?????
Endogenous chromatin structure
B
M1
M2
AB1380
YAC1
YAC2
YAC3
} YACs
C
Normalized nucleosome occupancy
Endogenous
YAC
KLLA0C03828g
-500
TSS
+1 kb
D
KLLA0C04411g
-500
TSS
+1 kb
E
F
K. lactis (3 YACs)
C
D
Endogenous
YAC

Strains	Description
BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0
AB1380	MATa ura3-52 trp1-289 lys2-1 ade2-1 can1-100 his5 ρ+ ψ+
K. lactis	CLIB 209
D. hansenii	NCYC 2572
K. waltii	NCYC 2644
YAC1	AB1380 + 128 kb YAC (*K. lactis* Chromosome F 872404~1000550)
YAC2	AB1380 + 143 kb YAC (*K. lactis* Chromosome C 339713~482935)
YAC3	AB1380 + 136 kb YAC (*K. lactis* Chromosome C 443175~578764)
YAC6	AB1380 + 115 kb YAC (*D. hansenii* Chromosome C 1165392~1280355)
YAC7	AB1380 + 216 kb YAC (*D. hansenii* Chromosome D 1148162~1364529)
YAC12	AB1380 + 120 kb YAC (*K. waltii* contig S0 133276~N/A*)
YAC14	AB1380 + 131 kb YAC (*K. waltii* contig S33 510773~641751)

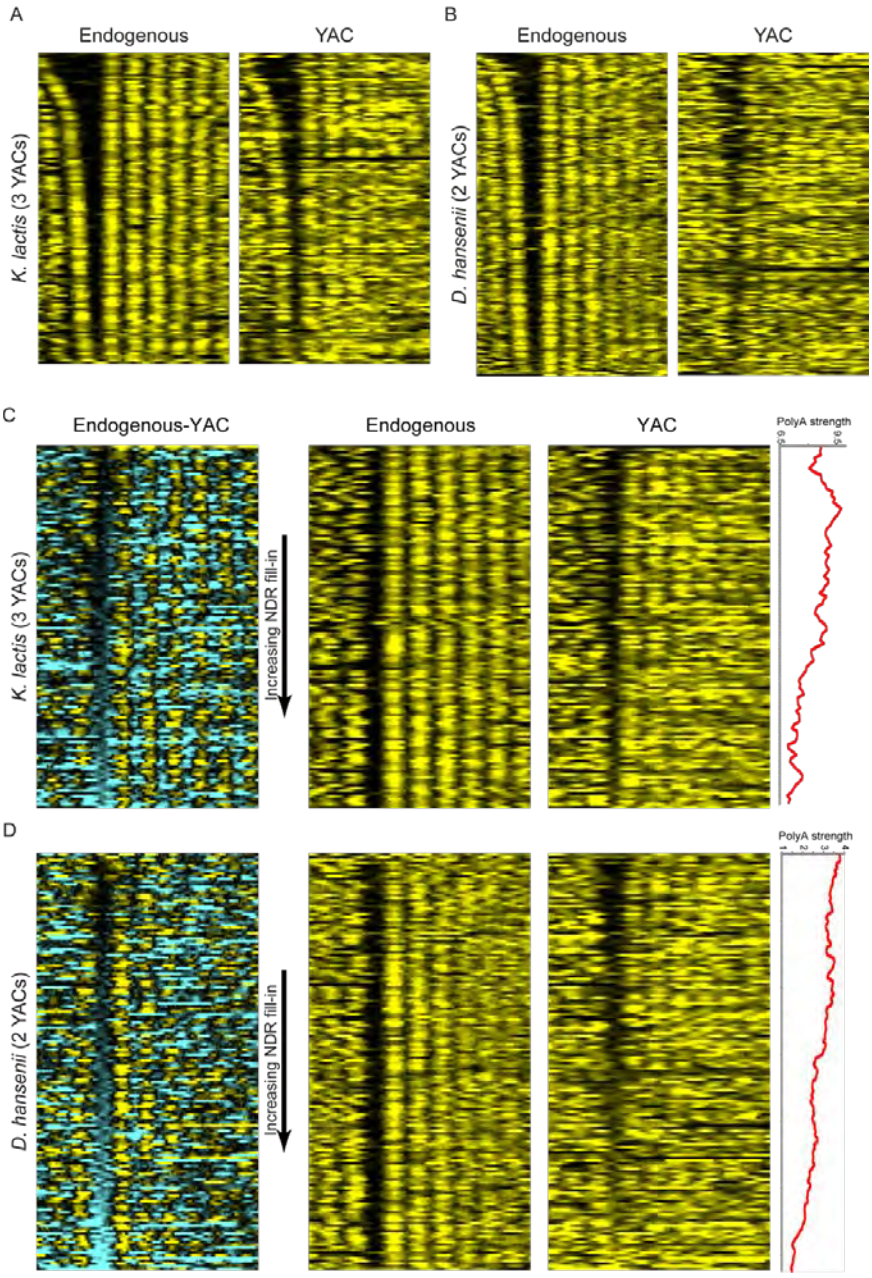
Table III. 1: Species and YAC strain list

Strains used in this study

*N/A: The coordinate is unable to be identified, due to incomplete *K.waltii* genome sequences.

Promoter NDRs are largely maintained in a foreign species in a manner strongly correlated with poly(dA:dT) tracts

The endogenous positions of promoter NDRs were largely maintained in the YACs (Figures III.1E-F) – 50% and 56% of *D. hansenii* and *K. lactis* NDRs, respectively, were located within 50 bp of their endogenous position, and for both species' sets of YACs only 13% of NDRs did not overlap the endogenous NDR at all. Furthermore, the extent of the NDR, which varies considerably among genes, is largely maintained in the YAC-containing strains. These data are consistent with the view that nucleosome depletion at fungal promoters is largely programmed by genomic sequence. However, the average extent of depletion over promoters is not as great in YACs as in wild-type (Figures III.1C-F, and Supplementary Figures III.1A-B), potentially as a consequence of the reduced expression of YAC genes (see below). This observation suggests that some of the depletion at promoters is not intrinsically determined by DNA sequence, consistent with the previous observation that nucleosome depletion at promoters is more pronounced *in vivo* than *in vitro* (Kaplan et al., 2009; Zhang et al., 2009; Zhang et al., 2011).



Supplementary Figure III.1: NDRs are generally better-maintained over sequence from *K. lactis* than over *D. hansenii* sequence

(**A-B**) Nucleosome mapping data for all genes from *K. lactis* (**A**) or *D. hansenii* (**B**) are shown for wild-type and YACs, as indicated. Genes are sorted by wild-type NDR width.

(**C**) NDR maintenance correlates with poly(dA:dT) elements. As in Figures III.2A-B, but for *K. lactis*.

(**D**) Identical to Figures III.2A-B, reproduced here for comparison between *D. hansenii* and *K. lactis*.

To further investigate the role for intrinsic sequence cues in establishing nucleosome depletion, we sorted genes by the difference in nucleosome occupancy over the proximal NDR between endogenous genes and YACs (Figure III.2A, Supplementary Figures III.1C-D). Notably, in both *K. lactis* and *D. hansenii* we observed very few genes with lower nucleosome occupancy in the YAC context, with the majority of promoters showing a range from little change to substantially increased nucleosome occupancy. Genes that maintained the same level of nucleosome depletion in the YAC as in wild-type were characterized by promoter sequences with greater numbers of long poly(dA:dT) elements (Figures III.2B-C**,** see Methods), consistent with the idea that these sequences intrinsically program nucleosome depletion in any context. Genes from *D. hansenii* tended to exhibit much less dramatic nucleosome depletion at promoters in the YAC context than in their endogenous context (Supplementary **Figure III.1B**). This is consistent with our prior observation (Tsankov et al., 2011; Tsankov et al., 2010) that *D. hansenii* promoters have fewer poly(dA:dT) sequences than most other organisms in the *Hemiascomycota* phylogeny (potentially due to their ecological niche in high salt environments) and with the hypothesis that *D. hansenii* promoters are more often established by *trans*-acting proteins such as General Regulatory Factors (GRFs). In this regard, promoters that gained substantial nucleosome occupancy when carried in the YAC sometimes, but not always, contained known binding motifs for transcription factors we previously (Tsankov et al., 2011; Tsankov et al., 2010) inferred to be GRFs in *D. hansenii* but not in *S.*

cerevisiae (Figure III.2D). Together, these results indicate that intrinsic sequence determinants (or conserved *trans*-acting factors) play a major role in generating nucleosome depletion at fungal promoters, and that poly(dA:dT) tracts are the primary DNA sequence determinant.

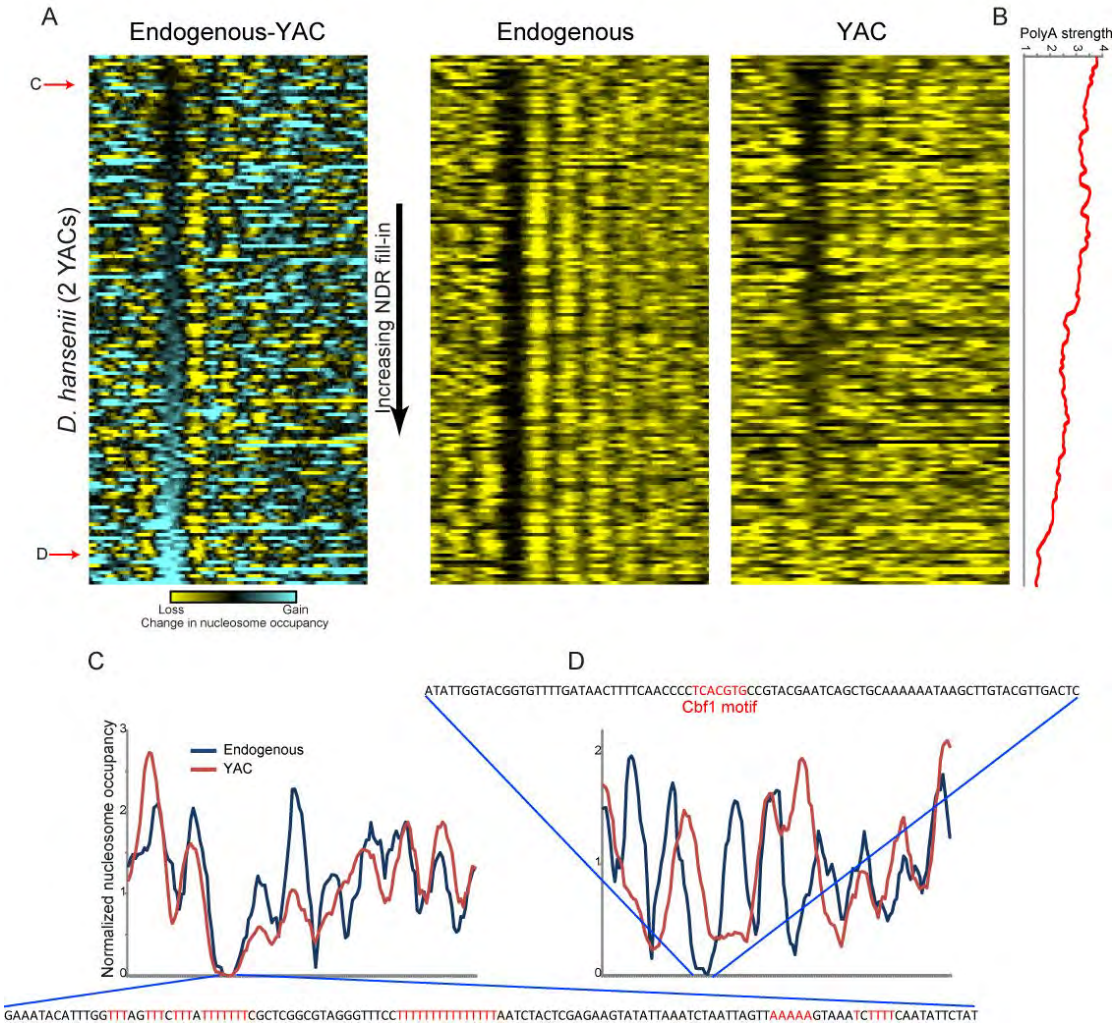
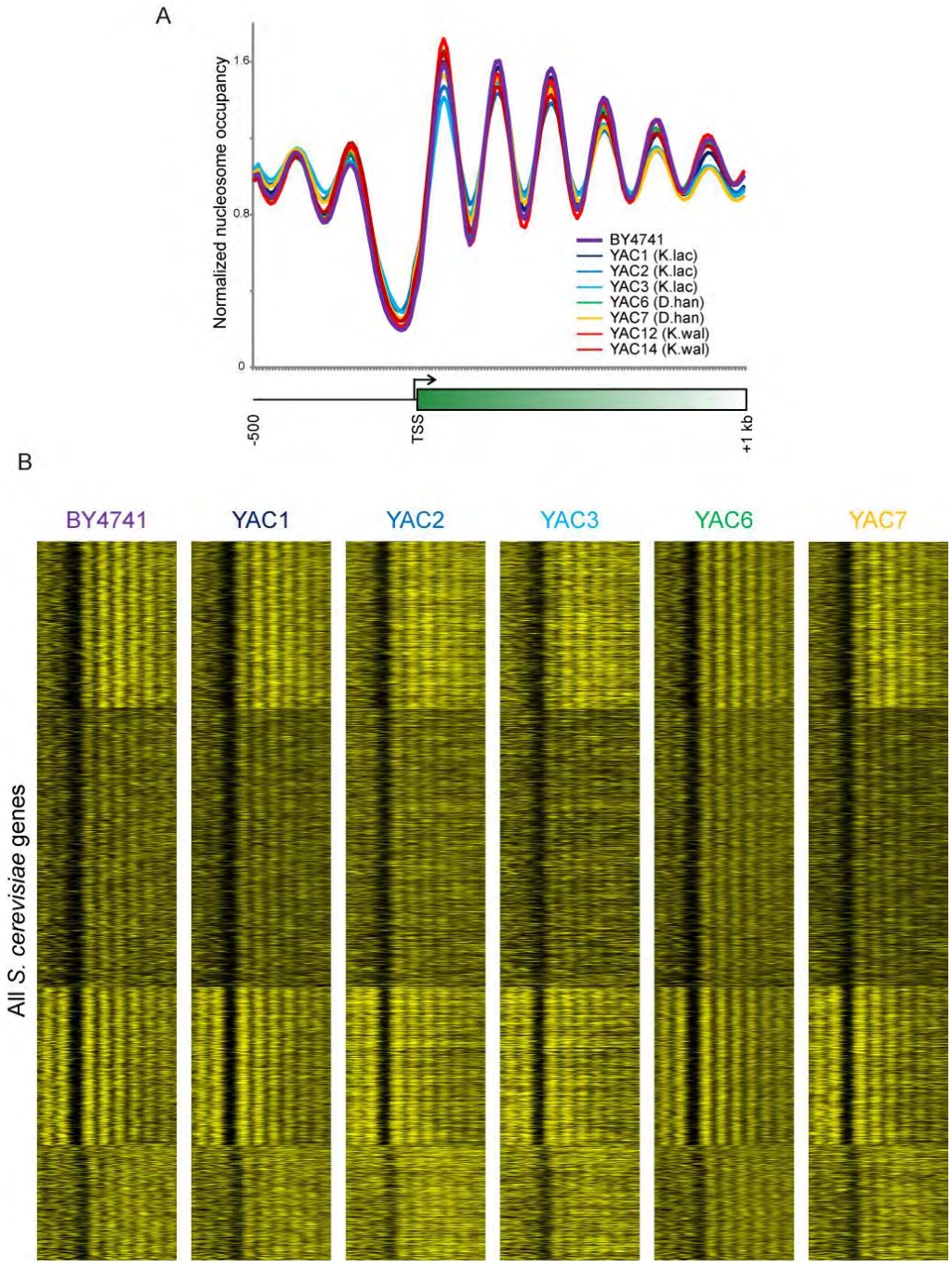


Figure III. 2 : Promoter nucleosome depletion is maintained over poly(dA:dT) elements.

(**A**) *D. hansenii* genes sorted by the extent of change in nucleosome occupancy over the NDR. Left panel shows differences in nucleosome occupancy between *D. hansenii* and YACs for 114 genes – blue indicates increased nucleosome occupancy in the YAC relative to endogenous context. Middle and right panels show nucleosome mapping data for endogenous *D. hansenii* sequences and for YACs, as indicated. (**B**) Strength of poly(dA:dT) element (Field et al., 2008; Tsankov et al., 2010) for genes, ordered as in (**A**). 40 gene running window average is shown. (**C**) An example of a gene with little change in nucleosome depletion between endogenous and YAC contexts. Sequence from this stable NDR contains multiple poly(dA:dT) elements, as indicated in red. (**D**) An example of a gene exhibiting dramatically increased nucleosome occupancy at the native NDR when carried on YAC. Here, this NDR includes few polyA elements, and carries a binding site for Cbf1, which has nucleosome-evicting activity in *D. hansenii* but not in *S. cerevisiae* (Tsankov et al., 2011; Tsankov et al., 2010).

Nucleosome positions differ markedly in the endogenous and YAC-containing strains

In contrast to the widespread but not universal conservation of NDRs, a given DNA sequence is generally packaged differently when carried in the endogenous species or in *S. cerevisiae*. Nucleosome positions change markedly in the YAC strains – the +1 nucleosome can be found near to (Figure III.1C), upstream (see below), or downstream (Figure III.1D) of its endogenous location, while nucleosomes farther downstream of the +1 occur farther and farther away from their endogenous locations. By definition, differences in nucleosome positioning of a given genomic region in the endogenous organism or in *S. cerevisiae* cannot be due to intrinsic DNA sequence, but rather *trans*-acting factor(s). These measured differences are not secondary to technical artifacts such as differences in MNase digestion, as we observe remarkably consistent results for *S. cerevisiae* genes for the various YAC datasets (Supplementary Figure III.2). Interestingly, the average deviation in chromatin structure between genes in their endogenous context and in the YAC was quite different for the three species studied – *K. lactis* genes appeared closest to their native structure when in YACs, whereas chromatin structure of *K. waltii* sequences in YACs appeared random with respect to genic structure (data not shown).



Supplementary Figure III.2 :Bulk chromatin is not affected in YAC-bearing strains

(**A**) Nucleosome sequencing data for all strains in this study was mapped to the *S. cerevisiae* genome, and data are averaged for all genes aligned by the +1 nucleosome.

(**B**) Data for all *S. cerevisiae* genes are shown for wild-type (BY4741) and 5 YAC-bearing strains. Genes are sorted by K means clustering (K=4) of BY4741 dataset.

Nucleosome spacing is determined by protein factors in the host organism, not DNA sequence

As observed by MNase cleavage of bulk chromatin, most nucleosomes in any given species are found in arrays in which the linker regions between adjacent nucleosomes are similar in size. Interestingly, nucleosome spacing is substantially different between *S. cerevisiae*, with an average internucleosomal distance of ~165 bp, and *K. lactis*, with an average spacing of ~178 bp (Heus et al., 1993; Tsankov et al., 2010). We therefore used our YAC dataset to assess whether nucleosome spacing over *K. lactis* genes is established by DNA sequence, or by the nuclear environment. As can be appreciated in Figures III.1C-D, nucleosome spacing appears shorter over *K. lactis* genes when carried in *S. cerevisiae*, relative to the endogenous spacing. Figure III.3A shows the average nucleosome data for all *K. lactis* genes present on one of the 3 YACs, aligned by the endogenous location of the +1 nucleosome. Average nucleosome spacing decreases when these genes are carried in YACs. The distribution underlying this average trend is quantified in Figure III.3B. Here, we called nucleosome positions (Weiner et al., 2010), then plotted the distribution of all internucleosomal distances as indicated. *K. lactis* genomic sequence in its endogenous context is packaged with nucleosomes occurring every 178 bp, whereas the same sequence in the *S. cerevisiae trans* environment exhibits ~165 bp nucleosome spacing, precisely the same spacing observed over native *S. cerevisiae* genes. Importantly, we observed no change in the spacing of *S.*

cerevisiae genes between wild type yeast and our YAC strains, indicating no artifactual effects on nucleosome spacing from, for example, MNase titration level (Figure III.3B, Supplementary Figure III.2).

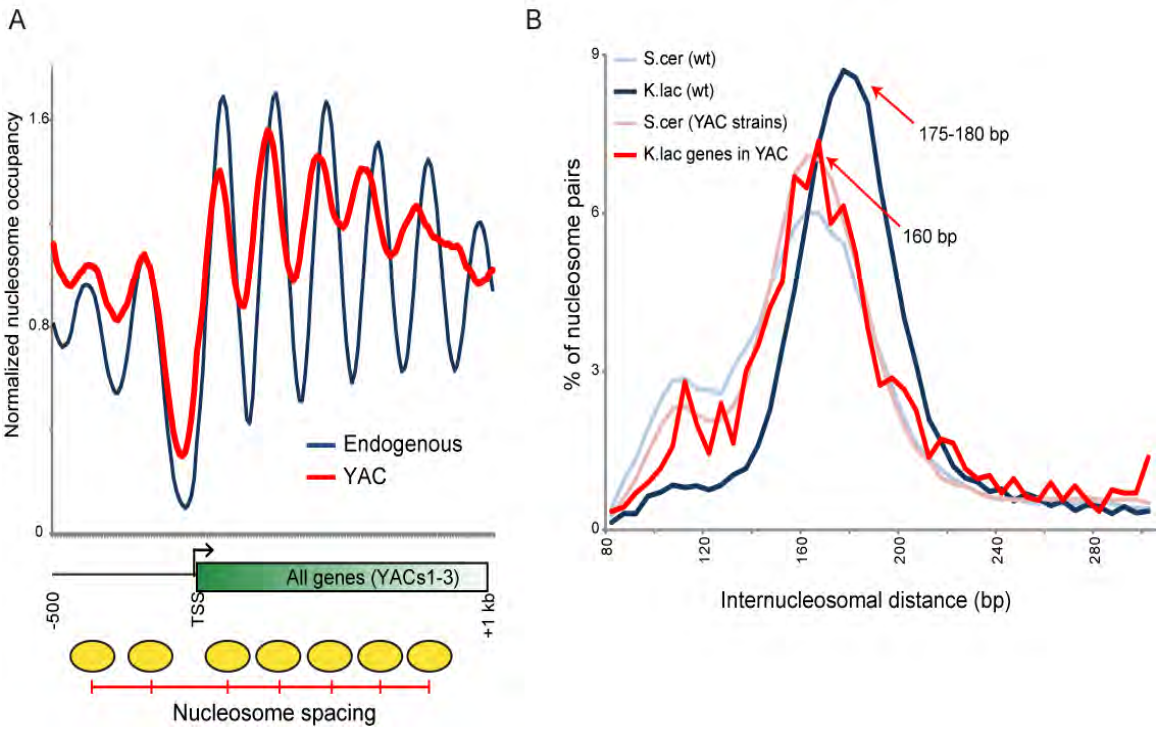


Figure III. 3: Nucleosome spacing is set in *trans*

(**A**) Averaged data for all *K. lactis* genes on YACs 1-3. Genes are aligned by the +1 nucleosome position as defined in Tsankov *et al.*, and data from either wild-type *K. lactis* or from the YAC strains are averaged for 184 genes, as indicated.

(**B**) *K. lactis* sequences adopt *S. cerevisiae* spacing when carried in *S. cerevisiae*. Nucleosome positions were called, and the distribution of all internucleosomal distances (center to center) is shown for 184 *K. lactis* genes from wild-type or in the YACs. Similar distributions for *S. cerevisiae* nucleosome positioning from wild-type and YAC-containing strains indicates that YACs do not perturb host chromatin state (See also Supplementary Figure III.2).

The primary role for proteins in determining internucleosomal spacing is not surprising, as different cell types in multicellular organisms (sharing identical genomes) can exhibit different nucleosome spacing (Van Holde, 1989). Importantly, the observation that internucleosomal spacing depends on protein factors means that the precise positions for the vast majority of nucleosomes are not determined by intrinsic DNA sequence.

The position of the +1 nucleosome is not determined by DNA sequence, but rather is mechanistically linked to transcriptional initiation

The claim that the +1 nucleosome is positioned by DNA sequence (Segal et al., 2006) has been subject to debate, not least because *in vitro* reconstitution experiments reveal no significant recovery of +1 nucleosome positioning. Alternatively, it has been proposed that the +1 nucleosome is positioned by either transcription factors such as Rap1, Abf1, and Reb1 (Kornberg and Stryer, 1988; Zhang et al., 2009; Zhang et al., 2011), NDRs (Mavrich et al., 2008; Yuan et al., 2005), or the preinitiation complex (Zhang et al., 2009).

Interestingly, although the average position of the +1 nucleosome is similar between YACs and the endogenous context for both *K. lactis* and *D. hansenii* genes (see Figure III.3A, Supplementary Figure III.1), examination of individual genes shows that +1 positioning is highly variable for the same sequence in two different nuclear environments. The distribution of +1 nucleosome shifts for genes carried on YACs was far more variable than the experimental variability measured using the background of *S. cerevisiae* genes –

while only 17% (14%-21% in different YAC strains) of *S. cerevisiae* +1 nucleosomes appeared > 20 bp apart between strains, 45% of *K. lactis* and 57% of *D. hansenii* +1 nucleosomes in the YAC strains shifted at least 20 bp from their endogenous location (Supplementary Figure III.3). +1 nucleosomes could shift in either direction on YACs (Figure III.4, Supplementary Figure III.3), although in *K. lactis* YACs these shifts were biased towards upstream shifts (Figure III.3A). Thus, our observations demonstrate that pronucleosomal sequences do not "program" the position of the +1 nucleosome *in vivo*.

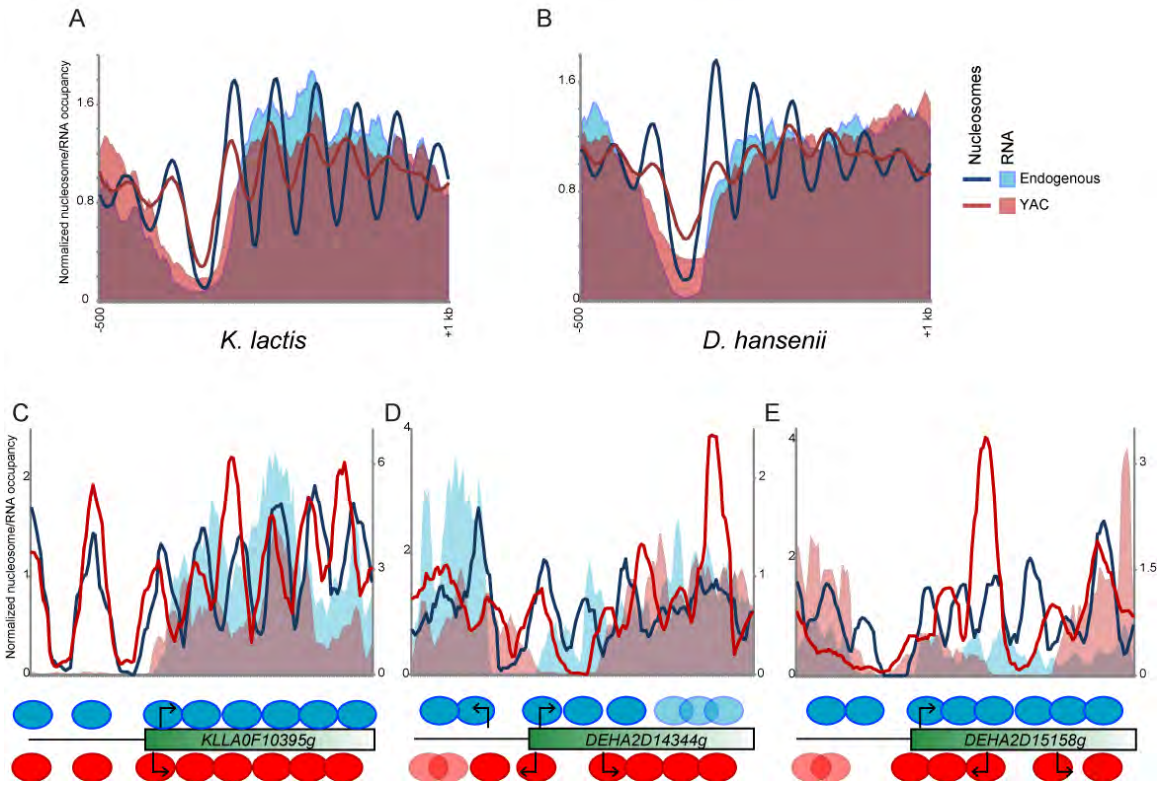
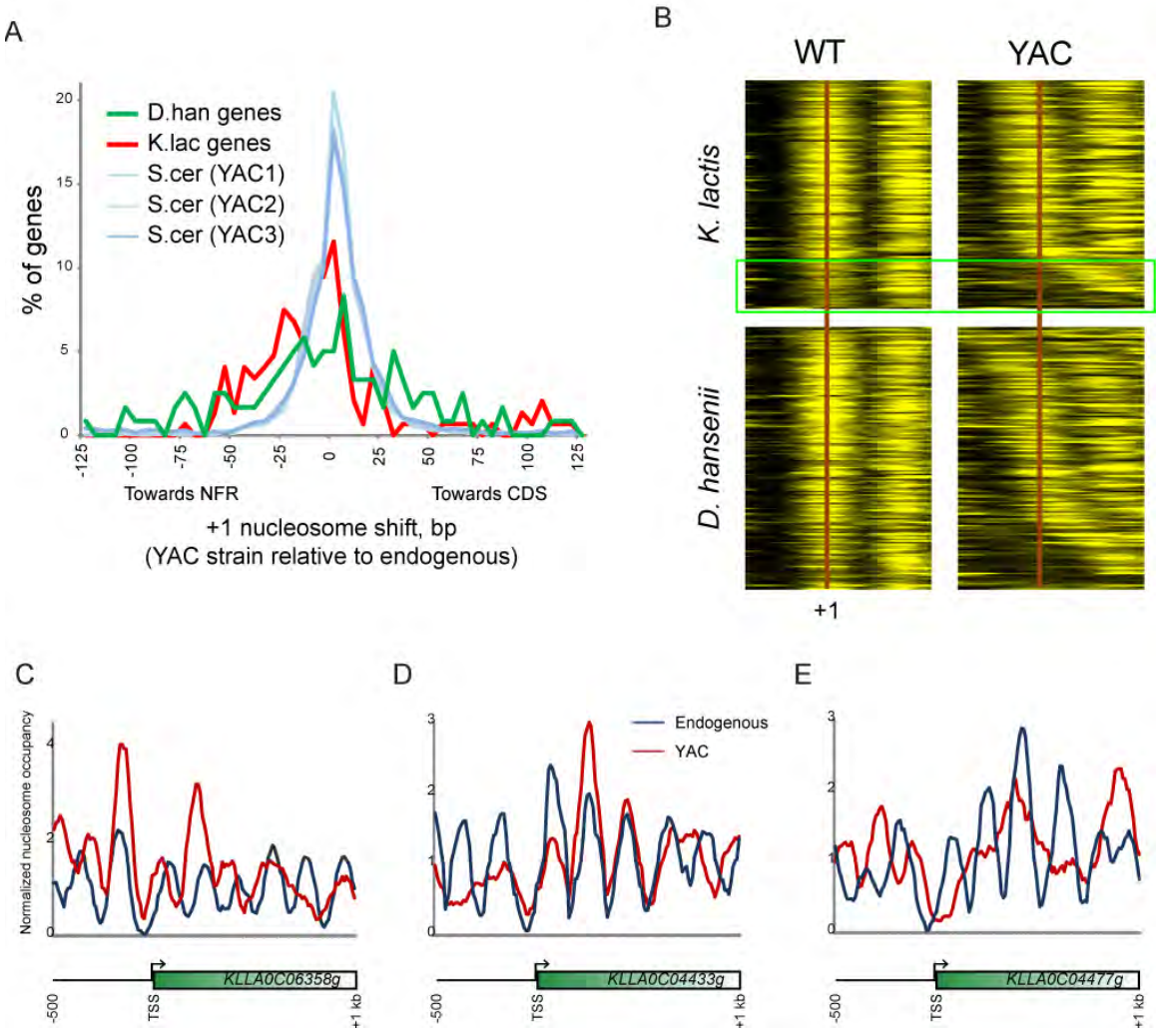


Figure III. 4: +1 nucleosome shifts associated with transcriptional changes.

(**A-B**) Nucleosome data and RNA-Seq data are shown for *K. lactis* and *D. hansenii* genes in wild-type and YACs, as indicated. RNA-Seq data for YAC-derived transcripts are normalized independently from *S. cerevisiae* transcripts here – see **Supplementary Figures III.4B-C** for data normalized genome-wide.

(**C-E**) Examples of +1 nucleosome shifts associated with changes in transcription. (**C**) shows a moderate upstream shift in a +1 nucleosome with a similar change in transcript length, while (**D-E**) show large scale NDR gain/loss with associated changes in transcription. Schematic interpretation of the nucleosome positioning for the endogenous gene is shown in blue above the rectangle, nucleosome positioning in the YAC is shown in red below the rectangle. Arrows indicate inferred TSSs (note that RNA-sequencing data are not strand-specific, but TFIIB mapping data support our inferred TSSs) – the furthest 5' RNA in (**E**), for example, derives from the upstream gene as opposed to a divergent promoter.



Supplementary Figure III.3: +1 nucleosome shifts associated with transcription.

(**A**) Distribution of shifts for +1 nucleosomes. Distributions are shown for the shifts between +1 positions in wild-type and YAC-bearing strains. The three *S. cerevisiae* distributions show changes in +1 positioning for *S. cerevisiae* genes in the indicated YACs, showing that technical variability or analytical variability do not account for nucleosome position changes for YAC-associated genes.

(**B**) Data for all *K. lactis* and *D. hansenii* genes, from wild-type and YACs. Data are shown from –100 to +300 bp relative to the +1 nucleosome upstream border (+1 nucleosome center indicated as a red line). Green box for *K. lactis* genes indicates a set of genes with low +1 nucleosome occupancy in wild-type, where the downstream shift in the YAC is likely due to failure to correctly call the low occupancy +1 nucleosome in the YAC.

(**C-E**) Examples of *K. lactis* genes exhibiting different +1 nucleosome shifts in the YAC context, including an upstream shift (**C**), an unchanged +1 (**D**), and a downstream shift (**E**).

The strong correspondence between +1 nucleosome positioning and transcriptional start sites in many species (Jiang and Pugh, 2009) led us to consider the hypothesis that changes in transcriptional activity might underlie the repositioning of the +1 nucleosomes (Zhang et al., 2009). We therefore carried out deep sequencing of RNA isolated from *D. hansenii*, *K. lactis*, and the *S. cerevisiae* YAC strains in this study, and carried out ChIP-Seq for TFIIB localization in the YAC-containing strains (a full analysis of these data will be published separately). Alignment of RNA-Seq data from wild-type strains with nucleosome mapping data confirmed prior predictions that the positioning of +1 nucleosomes with respect to a gene's transcription start site (TSS) varies between these species (Tirosh et al., 2007; Tsankov et al., 2010) – transcription begins further inside the +1 nucleosome in *K. lactis* than in *D. hansenii* (Supplementary Figure III.4A).

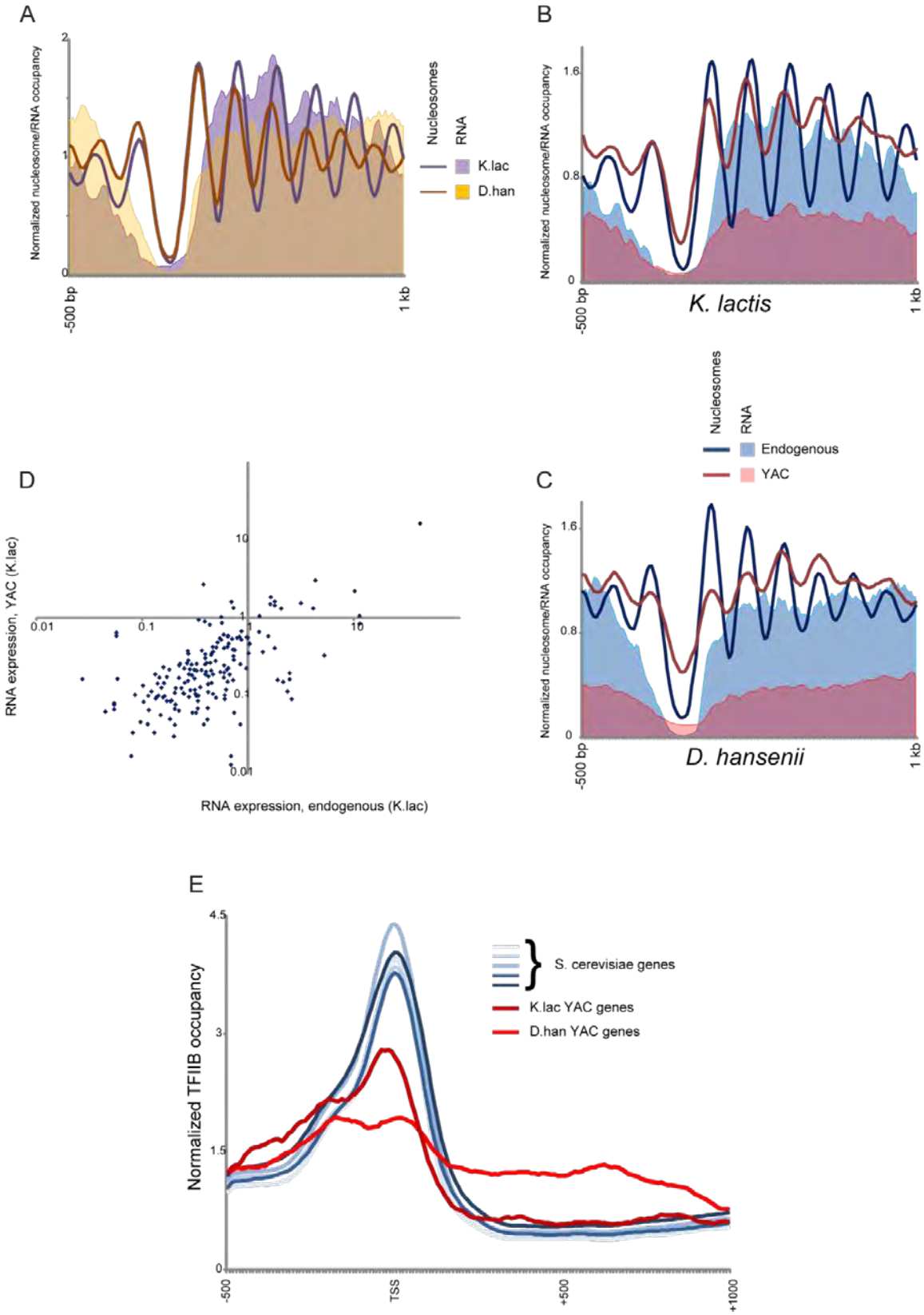
Supplementary Figure III.4: Comparison of RNA-Seq and MNase-Seq datasets.

(**A**) TSS positioning relative to +1 nucleosome. Averaged nucleosome data (solid lines) or RNA-Seq data (shaded area) for *K. lactis* and *D. hansenii* wild type cells, as indicated. All genes are aligned by +1 nucleosome position.

(**B-C**) Lower expression of genes on YACs relative to endogenous expression. For *K. lactis* (**B**) or *D. hansenii* (**C**), nucleosome data and RNA-Seq data are plotted as indicated for YAC-associated genes. For YAC-associated genes, RNA-Seq data are normalized to whole-genome RNA data (e.g. including *S. cerevisiae* genes, in contrast to the normalization in **Figure III.4**), with lower normalized abundance indicating that YAC-associated genes are expressed at lower levels than are endogenous genes, assuming similar RNA content of the various species.

(**D**) RNA abundance for *K. lactis* genes. Normalized RNA-Seq data from *K. lactis* (x axis) or YAC-bearing strains (y axis) is scatter-plotted with each point representing a single gene. mRNA abundance data are shown as reads per kb per million reads. Note good correlation between endogenous and YAC-based expression, indicating that differences between poorly and highly-expressed genes are maintained in a foreign nuclear environment.

(**E**) ChIP-Seq was carried out for TFIIB in the YAC-bearing strains. TSS-aligned data are shown for all *S. cerevisiae* genes in each strain, or for all YAC-based genes for either *K. lactis* or *D. hansenii*. TFIIB ChIP could not be carried out for the other organisms due to the limited cross-reactivity of our antibody. Note that foreign promoters continue to recruit TFIIB in *S. cerevisiae*, but that for both organisms TFIIB exhibits worse localization in the YAC than expected, due to divergence of regulatory information between species. This is especially true for *D. hansenii*.


A
Normalized nucleosome/RNA occupancy
2
1
0
-500 bp
1 kb
Nucleosomes
RNA
K.lac
D.han
B
Normalized nucleosome/RNA occupancy
1.6
0.8
0
-500 bp
1 kb
K. lactis
Nucleosomes
RNA
Endogenous
YAC
C
Normalized nucleosome/RNA occupancy
1.6
0.8
0
-500 bp
1 kb
D. hansenii
D
RNA expression, YAC (K.lac)
10
1
0.1
0.01
0.01
0.1
1
10
RNA expression, endogenous (K.lac)
E
Normalized TFIIB occupancy
4.5
3
1.5
0
-500
TSS
+500
+1000
S. cerevisiae genes
K.lac YAC genes
D.han YAC genes

Comparing endogenous to YAC-based gene expression, we found on average that genes on YACs were expressed at lower levels than host genes – average sequencing reads per kilobase of coding sequence for YACs was ~40% of the average value for endogenous RNAs – consistent with extensive promoter sequence divergence between species resulting in widespread misinterpretation of exogenous regulatory information by the *S. cerevisiae* transcriptional machinery (Supplementary Figures III.4B-C and E). In general, we found a good correlation between expression levels for genes in their endogenous genome versus expression from the YACs (Supplementary Figure III.4D) – genes expressed at high levels in *K. lactis* remained the most highly-expressed genes when carried on YACs, but were expressed at lower levels relative to *S. cerevisiae* genes. In *D. hansenii*, we also observed increased expression of intergenic regions in the YACs (Supplementary Figure III.4C and see below), again indicating evolutionary divergence in transcriptional control sequences (e.g. loss of transcriptional termination signals and/or gain of cryptic promoters).

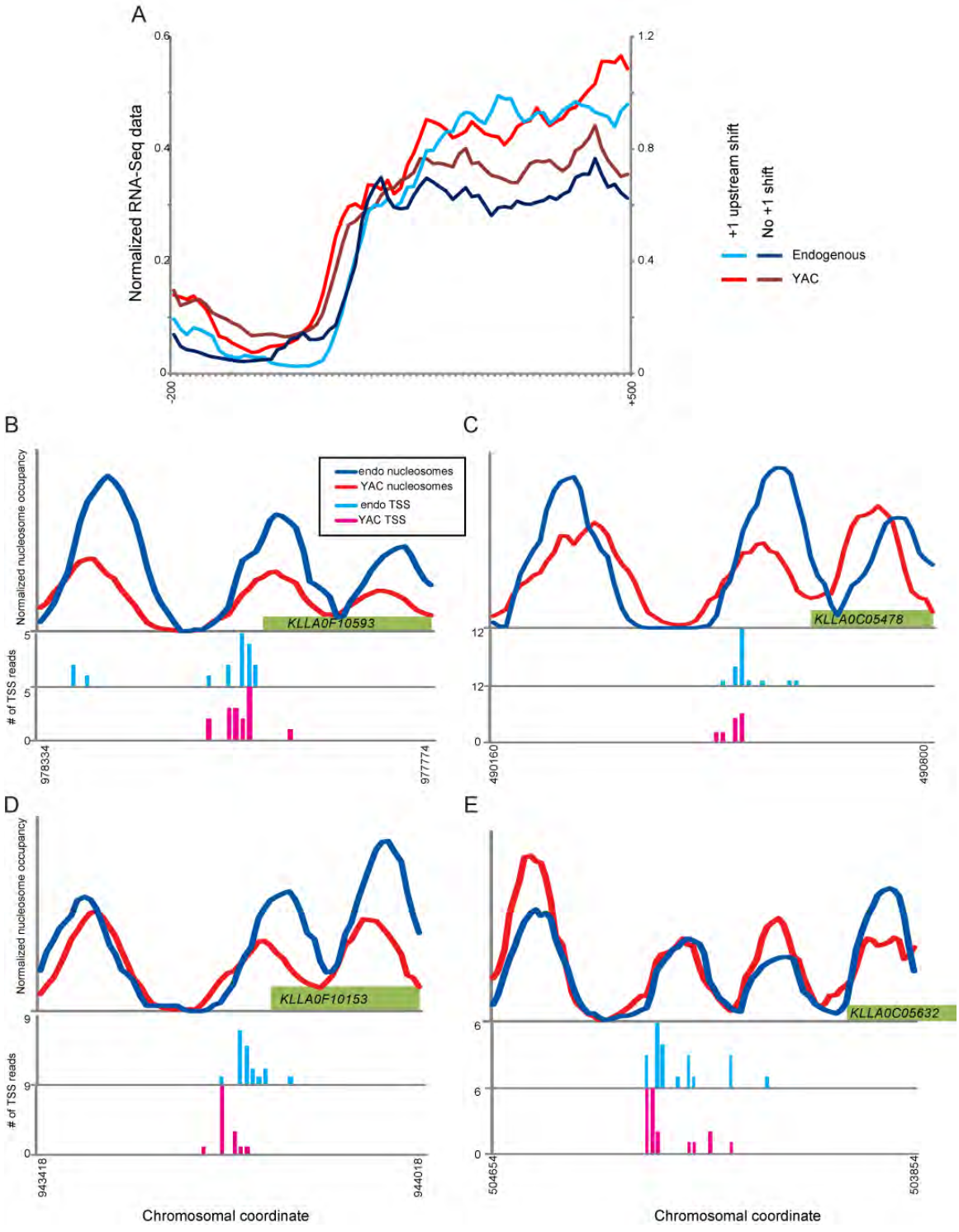
Consistent with a relationship between +1 nucleosome positioning and transcription start sites, we found that the 5' ends of RNAs in YACs shifted on average towards a *S. cerevisiae*-like location relative to the +1 nucleosome (Figures III.4A-B) – *K. lactis* RNAs started farther upstream in the YAC, whereas *D. hansenii* RNAs started farther downstream. Furthermore, +1 nucleosome shifts largely were accompanied by coherent shifts in inferred transcription start sites. These include ~90 examples such as that seen in Figure III.4C, in which

both RNA-Seq data and the +1 nucleosome for a given gene shift in the same direction. This is visualized in Supplementary Figure III.5A – RNA-Seq data for *K. lactis* genes exhibiting no +1 nucleosome shift, and for those exhibiting upstream +1 nucleosome shifts, is averaged for both endogenous RNA expression and YAC-based expression. Despite no average difference in 5' ends of transcripts between these two classes in the endogenous case, we find that genes exhibiting upstream shifts in the +1 nucleosome also showed more strongly 5'-shifted transcripts relative to genes without a +1 nucleosome shift, consistent with the idea that there is a mechanistic coupling between +1 nucleosome positioning and transcriptional initiation. Furthermore, we also used 5' RACE to more precisely map TSSs for 4 *K. lactis* genes in their endogenous context and in the YAC, confirming that +1 nucleosome shifts corresponded to shifts in the location of the TSS (Supplementary Figures III.5B-E). These observations provide functional evidence for a mechanistic linkage between nucleosome positioning and transcriptional initiation, although they do not establish the cause-and-effect relationship between these two processes.

Supplementary Figure III.5: Nucleosome positioning shifts are associated with shifts in TSS

(**A**) Averaged nucleosome mapping and RNA-Seq data are shown for all *K. lactis* genes exhibiting either no +1 nucleosome shift (less than 20 bp in either direction) in the YAC, or a 20 bp or more upstream shift in the YAC. Note that RNA-Seq data shift 5' in both cases when *K. lactis* genes are expressed in *S. cerevisiae*, but that genes with 5' nucleosome shifts exhibit a greater 5' shift, consistent with a constant distance being maintained between TSS and +1 nucleosome positioning. Note that genes with 3' shifts in the +1 nucleosome are not included, as these largely represent genes where the +1 nucleosome decreases occupancy and hence is miscalled (Supplementary Figure III.3B).

(**B-E**) TSS mapping for 4 individual genes. TSSs were mapped by 5' RACE (Methods), and individual clone locations are shown as indicated below the nucleosome mapping data. Note that for two genes with little +1 nucleosome position shift (**B**, **E**) there is little change in TSS, whereas the two genes with 5' shifts in the +1 nucleosome (**C**, **D**) also show upstream shifts in the TSS in the YAC.


A
Normalized RNA-Seq data
0.6
0.4
0.2
0
1.2
0.8
0.4
0
-200
+500
+1 upstream shift
No +1 shift
Endogenous
YAC
B
C
D
E
endo nucleosomes
YAC nucleosomes
endo TSS
YAC TSS
Normalized nucleosome occupancy
# of TSS reads
KLLA0F10593
KLLA0C05478
KLLA0F10153
KLLA0C05632
978334
977774
490160
490800
943418
944018
504654
503854
Chromosomal coordinate
Chromosomal coordinate

Generation of NDRs in foreign coding regions via fortuitous interactions of *S. cerevisiae* activator proteins

More dramatic cases of non-conserved nucleosome positioning are observed, particularly in *D. hansenii* -derived YACs, in which many novel NDRs arise in coding regions (Figures III.4D-E, Figures III.5A-B, Supplementary Figures III.6A-B). The sequences underlying these NDRs are not associated with poly(dA:dT) elements (analysis not shown), as expected given that they do not intrinsically form NDRs in their host genomic context (Figure III.5B, blue line). Interestingly, these novel NDRs are associated with TFIIB binding (Figures III.5C-D) and concomitant changes in RNA abundance (Figures III.4D-E, Figure III.5A, Supplementary Figure III.6A) indicating a wholesale functional change in which a coding sequence from one species (*D. hansenii*) is used as a promoter in a foreign species (*S. cerevisiae*). These NDRs are most likely determined by *S. cerevisiae* transcription factors that fortuitously recognize and functionally act on foreign DNA sequences that do not act as promoters in the native organism. In other words, DNA-binding transcriptional activator proteins recruit nucleosome-remodeling complexes to these fortuitously recognized sites, thereby evicting histones and generating an NDR. These NDRs are associated with varying levels of TFIIB and RNA transcripts, presumably depending on the quality of TATA elements and other core promoter sequences in the vicinity.

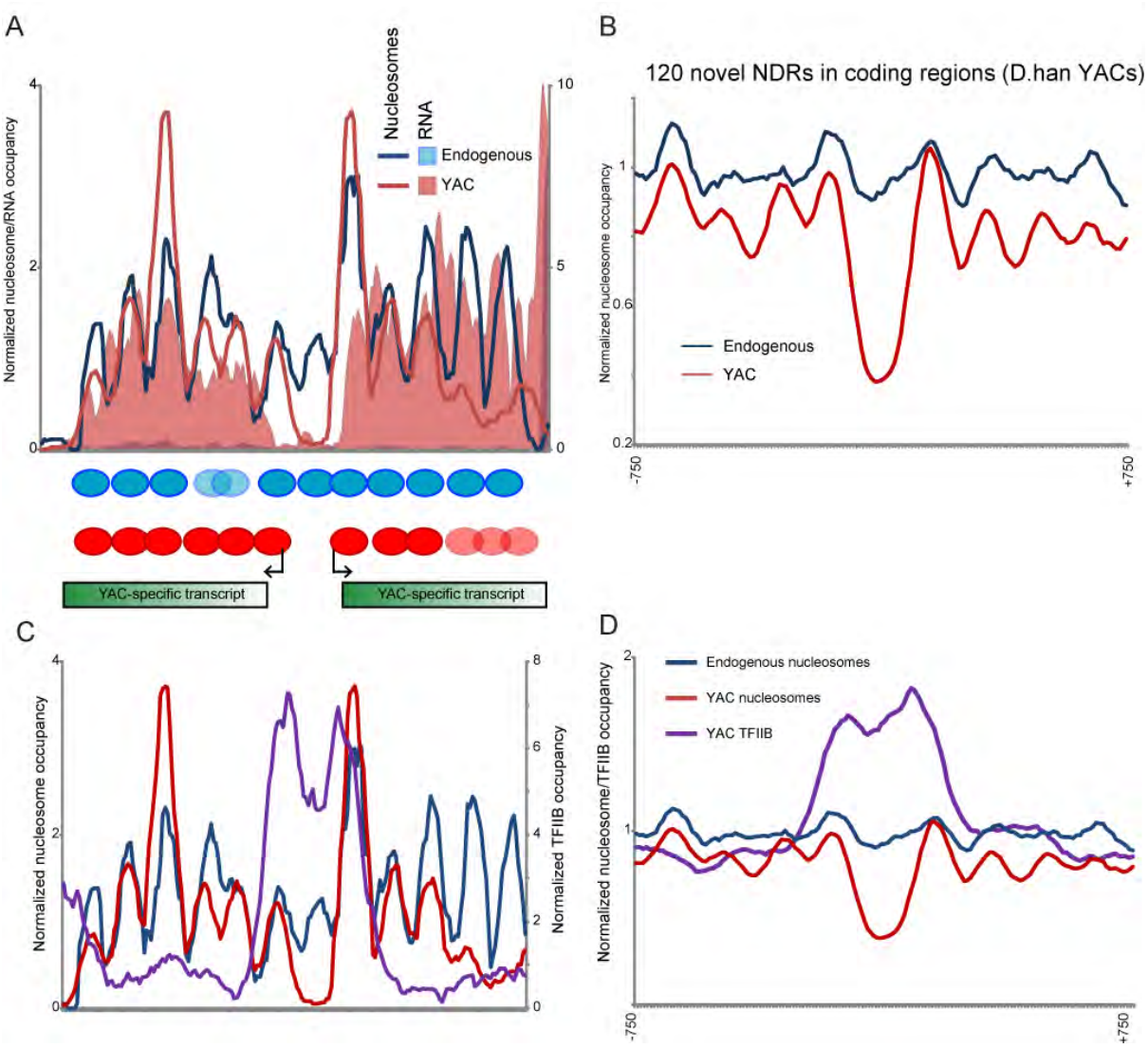
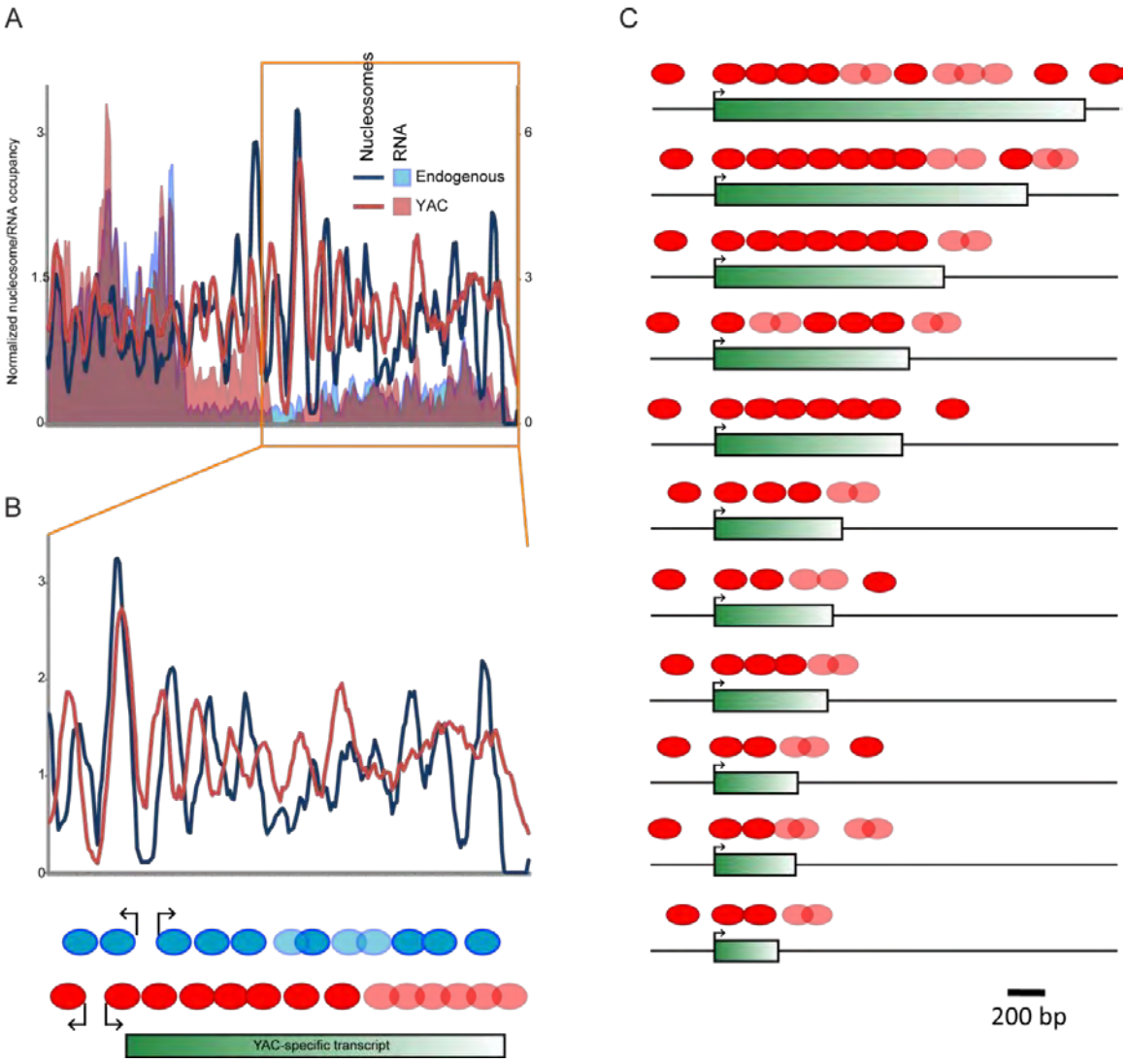


Figure III. 5: Characterization of novel NDRs in YACs.

(**A**) Example of a novel NDR that occurs only in the YAC but not in the native genome, and is associated with transcription. This new NDR occurs in the middle of a *D. hansenii* coding region, and is associated with two new shorter, divergent transcripts in the YAC context (data cover 2.2 kb of sequence). Note that nucleosome organization correlates with transcript length – rightmost transcript shows greater nucleosome positioning at the 5' end than at the 3' end of the transcript.

(**B**) Novel NDRs are generally associated with well-positioned +/-1 nucleosomes. Averaged data for 120 NDRs observed in *D. hansenii* YACs but not in the endogenous context, as indicated.

(**C-D**) Novel NDRs represent functional promoters. (**C**) shows TFIIB ChIP-Seq data from YAC-bearing strain for the genomic locus shown in (**A**), while (**D**) shows averaged data for all novel NDRs. Note that TFIIB localization in the endogenous context could not be obtained as our anti-TFIIB antibody does not recognize TFIIB from *D. hansenii*.



Supplementary Figure III 6: Examples of new NDRs in *D. hansenii* YACs.

(**A-B**) As in **Figure III.5**, another example of a novel NDR that occurs only in YACs, and is associated with transcription. (**A**) shows nucleosome and RNA data for 4 kb surrounding a YAC-specific NDR, with (**B**) showing nucleosome data only for the indicated region. Note increasing nucleosome fuzziness in the YAC nucleosome data at the 3' end of the novel transcript.

(**C**) Extent of positioned nucleosome array is linked to novel RNA transcript length. Schematic interpretation of the nucleosome positioning for RNA transcripts with different lengths, derived from novel coding region NDRs. RNA transcript is shown in green rectangle and nucleosome positioning is shown in solid red (well positioned nucleosome) and transparent light red (less positioned nucleosome) above the rectangle. Black arrows indicate inferred TSSs.

Fortuitous coding region NDRs are associated with typical nucleosome patterns

The existence of fortuitous and presumably evolutionarily meaningless promoters in *D. hansenii* coding regions in the context of *S. cerevisiae* cells makes it possible to determine the role of transcription in establishing the nucleosome positioning pattern. Strikingly, these coding region NDRs are associated with a typical nucleosome pattern of highly positioned +1 and -1 nucleosomes as well as progressively less positioned downstream nucleosomes (Figure III.5B). Thus in the absence of any intrinsic nucleosome-destabilizing sequences, transcription factors and associated co-factors are sufficient to generate a nucleosome positioning pattern that is very similar to the standard pattern at endogenous promoters. Furthermore, at such novel NDRs, the extent of the positioned array is linked to the length of the RNA transcript (Figure III.5A, Supplementary Figure III.6C), strongly suggesting a role for transcriptional elongation in the generation of the nucleosomal pattern. These results demonstrating a functional role for transcription-related events appear to conflict with the conclusion that nucleosome-remodeling complexes are sufficient to establish aspects of the nucleosome positioning pattern in the absence of transcription (Zhang et al., 2011). However, these observations are not mutually exclusive, and indeed are complementary as both mechanisms are likely to contribute to establishing the nucleosome pattern.

DISCUSSION

A functional evolutionary approach to address the determinants of molecular phenomena *in vivo*

Here, we used a functional evolutionary approach to systematically dissect the role for *cis*-acting sequence elements and *trans*-acting proteins in establishment of nucleosome positioning in fungi. This approach, which is based on species-specific differences in parameters of nucleosome positioning in a variety of yeast species (Tsankov et al., 2010), involves placing large segments of foreign yeast DNA in *S. cerevisiae* and comparing molecular properties in such strains with those in the native organism. In principle, non-conserved properties are determined by protein factors that are functionally distinct in the two species, whereas conserved properties are due either to DNA sequence or to conserved *trans*-acting regulators. The use of yeast artificial chromosomes to carry the foreign yeast DNA makes it possible to examine many genes at once, and hence to obtain information that is both statistically robust and permits one to identify many examples of new phenomenon. Furthermore, the ability to generate fortuitous functional events (e.g. the NDRs in *D. hansenii* coding regions) that do not occur in the native organisms makes it possible to address mechanistic questions in a manner that is, most likely, independent of evolutionary history. An extension of this approach should also permit one to identify factors responsible for the species-specific behavior, specifically by replacing a

candidate factor by its homolog in the foreign species and examining whether the pattern resembles that of the foreign species.

More generally, this functional evolutionary approach should allow for elucidating the determinants of other molecular phenomena that are broadly conserved but show species-specific differences. For example, more detailed analysis of the RNA generated in the YAC-containing strains with the corresponding endogenous yeast species should be informative of determinants of 5' and 3' end formation, splicing, and half-lives. As such, this approach combines the virtues of evolutionary comparison and classic functional genetic analysis.

A three-step model for establishing the nucleosome positioning pattern *in vivo*

Based on results presented here and elsewhere, we propose a three-step model (**Figure III.6**) for how nucleosome positioning is established in eukaryotic organisms. The first step involves the generation of an NDR, which can occur either by transcription factors and their recruited nucleosome remodeling complexes and/or by poly(dA:dT) sequences that intrinsically disfavor nucleosome formation. Even at poly(dA:dT)-containing promoters, it is likely that transcriptional machinery contributes to nucleosome depletion, as nucleosome depletion is more pronounced *in vivo* than *in vitro* (Kaplan et al., 2009; Zhang et al., 2009), and nucleosome-remodeling complexes enhance the depletion *in vitro* (Zhang et al., 2011). In this sense, intrinsic programming of NDRs represents a

specialized mechanism that is used frequently (*S. cerevisiae*), moderately (*D. hansenii*), or rarely (*D. melanogaster*), depending on the species.

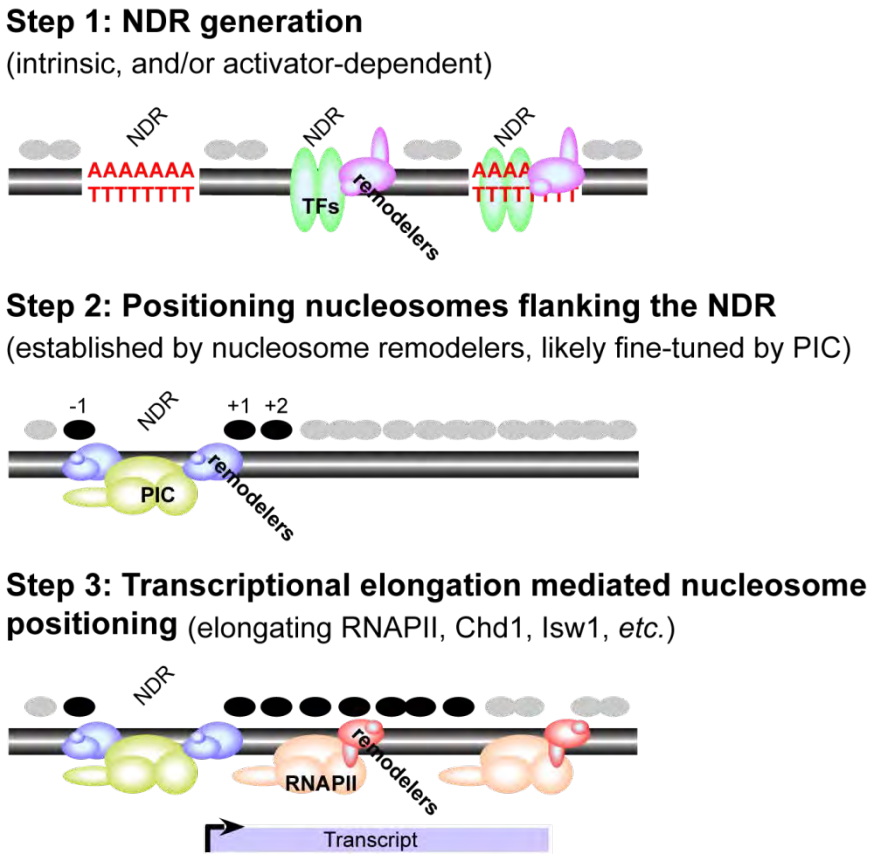
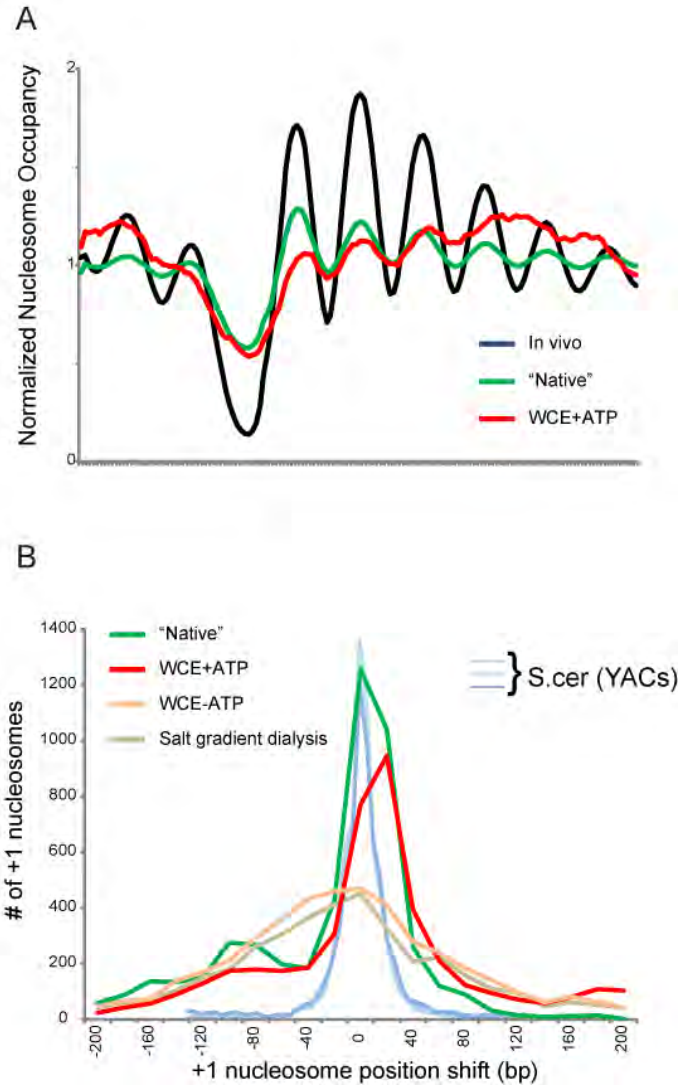


Figure III. 6: Three-step model for establishment of nucleosome positioning *in vivo*.

A unifying three-step model for how nucleosome positioning pattern is generated in eukaryotic organisms. The first step is the generation of an NDR, either by poly(dA:dT) elements and/or by transcription factors and their recruited nucleosome remodeling complexes. In the second step, nucleosome-remodeling complexes recognize the NDRs and generate highly positioned nucleosomes flanking the NDR; and the RNA polymerase II preinitiation complex fine-tunes the position of the +1 nucleosome. In the final step, positioning of the more downstream nucleosomes depends on transcriptional elongation, and the recruitment of nucleosome-remodeling activities and histone chaperones by the elongating RNA polymerase II machinery.

In the second step, nucleosome-remodeling complexes recognize the NDRs and generate highly positioned nucleosomes flanking the NDR. Strong positioning could, in principle, arise simply from the boundary of the NDR and/or from sequence preferences of the nucleosome remodelers. Indeed, it has been argued that this step does not require transcription factors or transcription *per se* (Zhang et al., 2011), although it is important to note that there is overall poor correspondence between +1 nucleosome positioning observed *in vivo* and that recapitulated using ATP-dependent extracts in the absence of transcription (Zhang et al., 2011) (Supplementary Figure III.7). In this regard, Zhang *et al.* compared nucleosome positioning generated by ATP-dependent extracts with the nucleosome positions measured from yeast lysed without crosslinking and allowed to redistribute prior to crosslinking. Indeed, we find mediocre correspondence between the "native" nucleosome positions from Zhang *et al.* and true *in vivo* nucleosome positions generated from crosslinked yeast (Supplementary Figure III.7), so the ability of whole cell extracts to recover these "native" positions in the absence of transcription does not have any bearing on the question of whether *in vivo* positioning is influenced by transcription prior to lysis of cells. Although nucleosome remodelers can generate somewhat positioned nucleosomes flanking the NDR and unquestionably perform far better than salt dialysis, they apparently are insufficient to generate the precise *in vivo* nucleosome positions, particularly for the +1 nucleosome (Supplementary Figure III.7).



Supplementary Figure III. 7: Yeast whole cell extracts poorly position nucleosomes

(**A**) Nucleosome mapping data from intact yeast ("*in vivo*"), from lysed yeast equilibrated prior to crosslinking ("native"), or from yeast genomic DNA incubated with yeast whole cell extract and ATP (WCE+ATP) *in vitro* (Zhang et al., 2011). Yeast whole cell extract performs significantly better than salt gradient dialysis, as reported, but the nucleosome positions recovered nonetheless do not precisely match nucleosome positions recovered from intact yeast – compare "native" and *in vivo* positioning.

(**B**) +1 nucleosome positioning in "native" yeast extracts exhibits systematic deviation from *in vivo* positioning. +1 nucleosome positions were called, and distance from +1 nucleosome positions *in vivo* to the positions recovered in various datasets is shown as a histogram. As a comparison for technical variation, we show data for +1 positioning variability in our YAC strains (to keep datasets on the same y axis we used a 5 bp bin size for YACs rather than the 10 bp used for other datasets).

Here, the strong, and species-specific, spacing relationship between the +1 nucleosome and mRNA start site that is observed both in the native and YAC strains indicates that there is a mechanistic connection between transcriptional initiation and the location of the +1 nucleosome. Given the strong *in vivo* positioning of both the preinitiation complex and the +1 nucleosome, a spacing relationship between these two entities requires that at least one of these is anchored to a specific location, thereby permitting a defined location for the second entity. As discussed above, nucleosome remodeling complexes alone are insufficient to generate proper positioning of the +1 nucleosome, and hence sequence and nucleosome remodelers are insufficient to provide an anchor. In contrast, preinitiation complexes bound at core promoters are clearly sufficient to provide an anchor, with the location of the TBP bound to the TATA element or TATA-related sequence being the major determinant of the anchor point. From these considerations, and our finding that the TSS to +1 distance in YACs shifts to the *S. cerevisiae* spacing (**Figure III.5 and Supplementary Figure III.5**), we suggest that the preinitiation complex plays a role in fine-tuning the position of the +1 nucleosome.

In the third step, positioning of downstream nucleosomes, with progressively less positioned nucleosomes downstream within the gene, depends on transcriptional elongation, and hence recruitment of nucleosome-remodeling activities and histone chaperones by the elongating RNA polymerase II machinery. This elongation-dependent step explains why nucleosome-

remodeling complexes, though capable of weakly positioning nucleosomes flanking the NDR, are unable to position more downstream nucleosomes (Zhang et al., 2011). Conversely, yeast mutant strains lacking nucleosome-remodeling complexes (Chd1 and Isw1) that are recruited to coding regions by elongating RNA polymerase show drastically reduced positioning of downstream nucleosomes, but relatively normal positioning of the +1 and +2 nucleosomes (Gkikopoulos et al., 2011). Finally, a transcription-based step nicely helps to explain why nucleosome arrays occur largely in the transcribed direction even though highly positioned nucleosomes can occur both at the +1 and -1 position, as well as the curious observation that the decay of nucleosome positioning towards the center of genes displays a 5'/3' asymmetry (Vaillant et al., 2010); both of these observations are inconsistent with a pure packing-based model.

The above model can explain why the general pattern of nucleosome positioning is highly conserved among eukaryotes, yet shows species-specific differences in various aspects of chromatin structure. These species-specific differences reflect the relative utilization of poly(dA:dT) sequences and hence intrinsic histone-DNA interactions, as well as differences in the enzymatic and recruitment properties of the nucleosome remodelers.

METHODS

Growth Conditions

All cultures were grown in medium containing: SC –Tryptophan –Uracil (Sunrise Sciences) (0.2%), Yeast extract (1.5%), Peptone (1%), Dextrose (2%), and Adenine (0.01%), as previously described (Tsankov et al., 2010).

Preparation of YACs

Yeast chromosomal DNA was prepared in InCert agarose blocks (LONZA), with a final cell concentration of 2 X 10^9 cells/ml. Agarose blocks with intact chromosomal DNA were subjected to EcoRI partial digestion with a titrated Mg^{2+} concentration, followed by size fractionation using pulsed field gel electrophoresis (PFGE). ~100-200 kb partially-digested DNA fragments were excised from the gel. YAC vector pYAC4 was purified by successive CsCl gradient ultracentrifugation and digested with BamHI and EcoRI, followed by calf intestine alkaline phosphatase treatment. Digested pYAC4 and partially-digested yeast chromosomal fragments were ligated by T4 DNA ligase in agarose blocks. Prior to YAC transformation, ligated DNA was size-fractionated again by PFGE and DNA larger than 100 kb was excised from the gel. The excised gel slice was further digested with β-agarase and ligated DNA was transformed into *S. cerevisiae* host cells (AB1380), using either spheroplast transformation protocol or standard yeast LiCl transformation method.

Validation of YAC-bearing strains

Transformants with red color, which survived double selection (Ura^+/Trp^+) on AHC plates, were collected for validation. Chromosomal DNA of candidate strains was prepared in agarose blocks and resolved by PFGE (**Figure III.1B**). Strains with desired YAC bands were selected, and terminal sequences from selected YAC clones were isolated and confirmed with DNA sequencing analysis (Riley et al., 1990).

Nucleosome Isolation and Illumina Deep Sequencing

Micrococcal nuclease (MNase) digestions were performed as previously described (Yuan et al., 2005). Briefly, 450mL cultures were grown to OD_{600} of ~0.5 at 30°C, 220rpm. Cultures were fixed for 30 minutes at 30°C with 1.85% formaldehyde, then spheroplasted with 10 mg zymolyase (Cape Cod Associates) for 45 minutes at 30°C. Spheroplasts were subjected to 20 minutes of MNase digestion, and DNA was purified. MNase titrations were selected to obtain largely mononucleosomal DNA with some di- and tri-nucleosomal DNA apparent. Mononucleosomal DNA was gel purified (BioRad Freeze N' Squeeze) and used to create a library for deep sequencing on the Solexa 1G Genome Analyzer, as previously performed Fast Link DNA Ligation Kit, Epicentre LK6201H) to Illumina genomic adapters, followed by a final PCR with a size-selecting gel purification (BioRad Freeze N' Squeeze 732-6166).

Data Normalization and Nucleosome TSS Alignments

Reads from deep sequencing were mapped back to the relevant hybrid genome (*S. cerevisiae* plus the relevant species' chromosome), using blat. Uniquely mapping reads that had fewer than three mismatches were kept for analysis. Reads were extended by the cross correlations between those from the Watson and Crick strands, to create nucleosome peaks. Read count numbers were normalized to one by dividing each base read count by the genome-wide average read count per base. Gene alignments were carried out using the endogenous boundary of the +1 nucleosome (Tsankov et al., 2010). RNA-seq data was treated similarly but without extending reads. RNA abundance for YAC-based transcripts was, on average, ~30-40% (in reads per kb per million reads) of the RNA abundance of endogenous *S. cerevisiae* transcripts.

Nucleosome Calls

Template Filtering (Weiner et al., 2010) was used to call the locations of nucleosomes.

5' RACE

Trizol (Invitrogen) extracted RNA was enriched for mRNA on polyT magnetic beads (NEB S1419S). Calf Intestinal Phosphatase (NEB M0290L) removed all phosphates prior to the hydrolysis of the mRNA cap to a phosphate with Tobacco Acid Pyrophosphatase (Epicentre T19250). An oligo was ligated to the 5' end of the mRNA (T4 RNA Ligase, NEB) and the RNA was reverse transcribed

(SuperScript III Reverse Transcriptase, Invitrogen) with a tailed random hexamer. The cDNA was amplified with a low cycle PCR (Phusion, NEB) using primers matching the sequences added in the ligation and reverse transcription. A gene specific PCR amplified the transcription start site sequence, which was cloned (StrataClone, Agilent) and sequenced.

CHAPTER IV: General Regulatory Factors play a role in promoter nucleosome depletion.

Introduction

The packaging of DNA into chromatin provides a layer of gene regulation. This is in part because nucleosomes restrict access to transcription factor binding sites. The regulation of chromatin structure has classically been studied at the PHO5 and GAL1/10 loci, where gene activation by activator binding requires and induces chromatin reorganization. The repression of these model genes is dependent upon nucleosomes, as activation regardless of the environmental context is achieved during histone-depletion (Han and Grunstein, 1988; Han et al., 1988b). While activator binding can compete away histone proteins *in vitro* (Workman and Kingston, 1992), it does not always appear to be a simple case of competition between activators and histones *in vivo* (Svaren et al., 1994; Workman et al., 1991). The specific architecture of nucleosome and transcription factor binding sites can lead to a plethora of regulatory effects. For instance, the presence of Pho4 binding sites in nucleosomal DNA contributes to the extent of activation, while those unobstructed by nucleosomes determine the signaling threshold of activation (Lam et al., 2008). Furthermore, artificially moving a nucleosome to its normal post-induction position allows for binding of TBP and relaxes the induction requirements for the virally induced IFNβ gene (Lomvardas and Thanos, 2002). Hence, the packaging of DNA into chromatin with regards to

promoter elements can influence gene expression by restricting access to binding sites.

Genome-wide studies have observed functional transcription factor binding sites to be enriched in nucleosome depleted regions (Lee et al., 2007). A simple explanation is that the binding of transcription factors would abrogate the binding of nucleosomes and vice versa. The ability to interfere with nucleosome positioning and thus induce gene expression has been seen to differ amongst transcription factors; in particular Gal4 has been seen to interrupt chromatin structure and induce reporter gene expression, while Gcn4 requires an additional Rap1 site (Yu and Morse, 1999). The factors with the greatest ability to disrupt nucleosome binding across main genomic loci have been classed as General Regulatory Factors (GRFs). These GRFs have often been seen to synergize with transcription factors to aid in their activity by disrupting nucleosome positioning to promote transcription factor binding (Buchman and Kornberg, 1990; Chasman et al., 1990). In some cases, GRFs have been seen to recruit chromatin remodelers to disrupt chromatin structure (Hartley and Madhani, 2009). Despite their similar and sometimes interchangeable effects, GRFs appear to work by different mechanisms to achieve the same goal (Yarragudi et al., 2004). Given their role in helping to deplete nucleosomes, GRF binding sites can be predicted in *S.cerevisiae* and other species (Kaplan et al., 2009; Tsankov et al., 2011) based on the disparity between *in vitro* and *in vivo* nucleosome depletion over specific sequences.

In order to examine the contributions of transcription and general regulatory factor binding to nucleosome depletion *in vivo*, we exploited the differences in nucleosome occupancy on the same sequence when in its genomic *Debaryomyces hansenii* context and on a yeast artificial chromosome in *S.cerevisiae*. We have previously discovered several nucleosome depleted regions at *D.hansenii* promoters that gain nucleosome occupancy in the *S.cerevisiae* environment, as well as several areas on *D.hansenii* yeast artificial chromosomes that lose nucleosome occupancy in the *S.cerevisiae* context. These could be due to the loss and gain of transcription factor binding, respectively, depending on the TFs present in the species. In order to systematically map transcription factor footprints as well as nucleosomes, we performed paired-end MNase-seq without a mononucleosomal size selection, as was pioneered in the Henikoff lab (Henikoff et al., 2011).

Results

Using paired-end sequencing of the entire product of a micrococcal nuclease digestion, we are able to map MNase footprints from approximately 20bp to mono- and di- nucleosomal sizes. Nucleosome depleted regions tend to be enriched for smaller protected fragments (Figure IV.1A&B), which presumably represent protection by transcription factors at the promoter. In some cases, multiple TF footprints map to nucleosome depleted regions (Figure IV.1A). The enrichment of TF footprints is reproducible from MNase digestions in the wild

type *S.cerevisiae* laboratory strain to the *S.cerevisiae* background of the *D.hansenii* yeast artificial chromosome strains. *D.hansenii* nucleosome depleted regions are also enriched for TF footprints (Figure IV.1B, right panel). However, the *D.hansenii* promoters that gain nucleosome occupancy in the *S.cerevisiae* environment lose this footprinting (Figure IV.2A), which is consistent with the idea that transcription factors present only in *D. hansenii* prevent the binding of histones. In order to see if the converse is true, we mapped TF footprints surrounding the newly depleted regions on *D.hansenii* sequence in the yeast artificial chromosome strains. The new NDRs do indeed coincide with the gain of TF footprints (Figure IV.2B). This is consistent with either the binding of TFs in the *S.cerevisiae* environment generating nucleosome depletion or the loss of nucleosome occupancy allowing for binding of TFs.

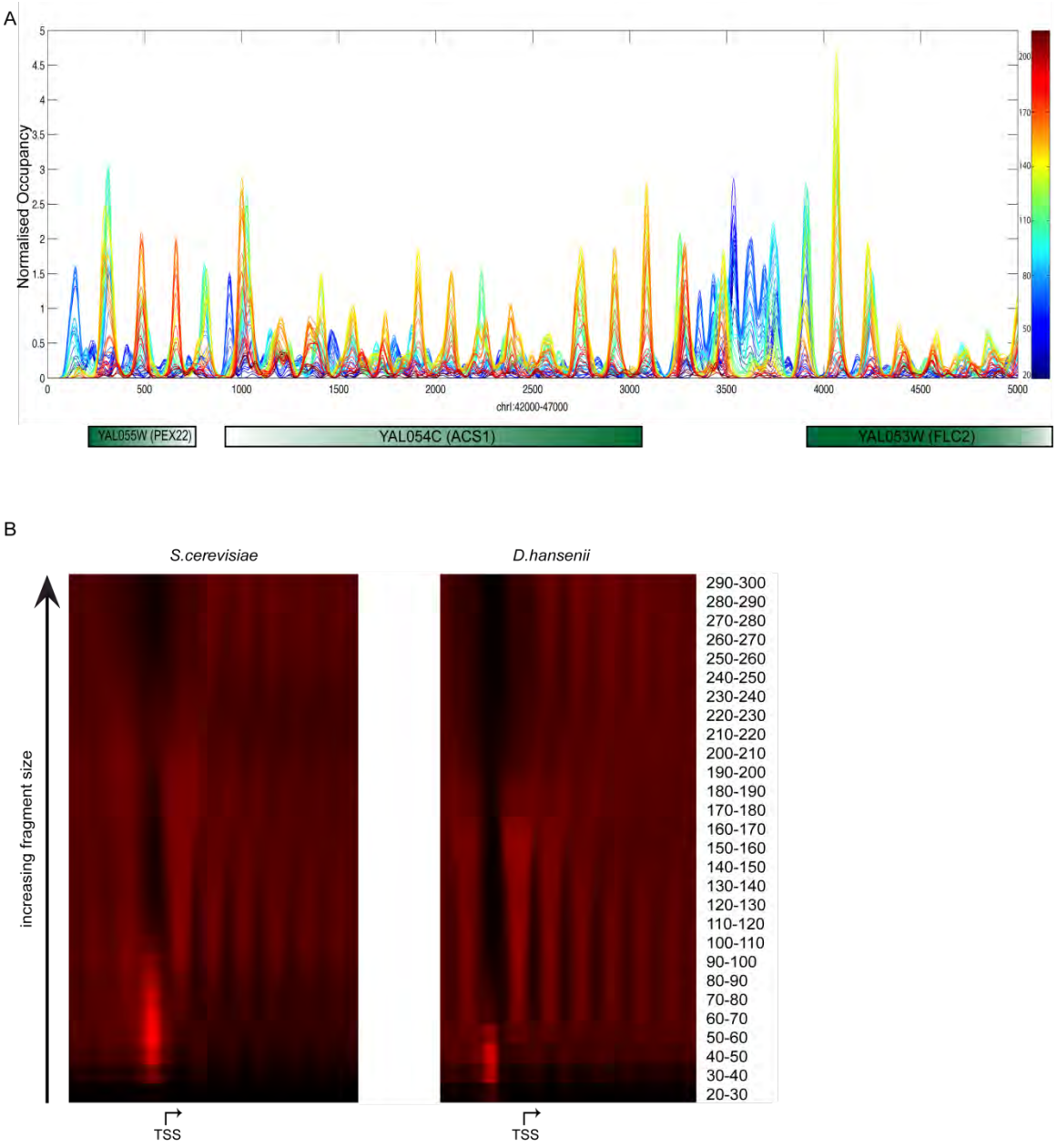


Figure IV. 1: Enrichment of transcription factor footprints in NDRs.

(A) Genome snapshot of MNase protected fragments in 10bp slices show enrichment of TF-sized footprints (sometimes multiple) in NDRs. Generated by Assaf Weiner in the Friedman lab.

(B) TSS aligned averages scaling from 20 to 300 base pairs in 10bp slices show average TF occupancy at promoters.

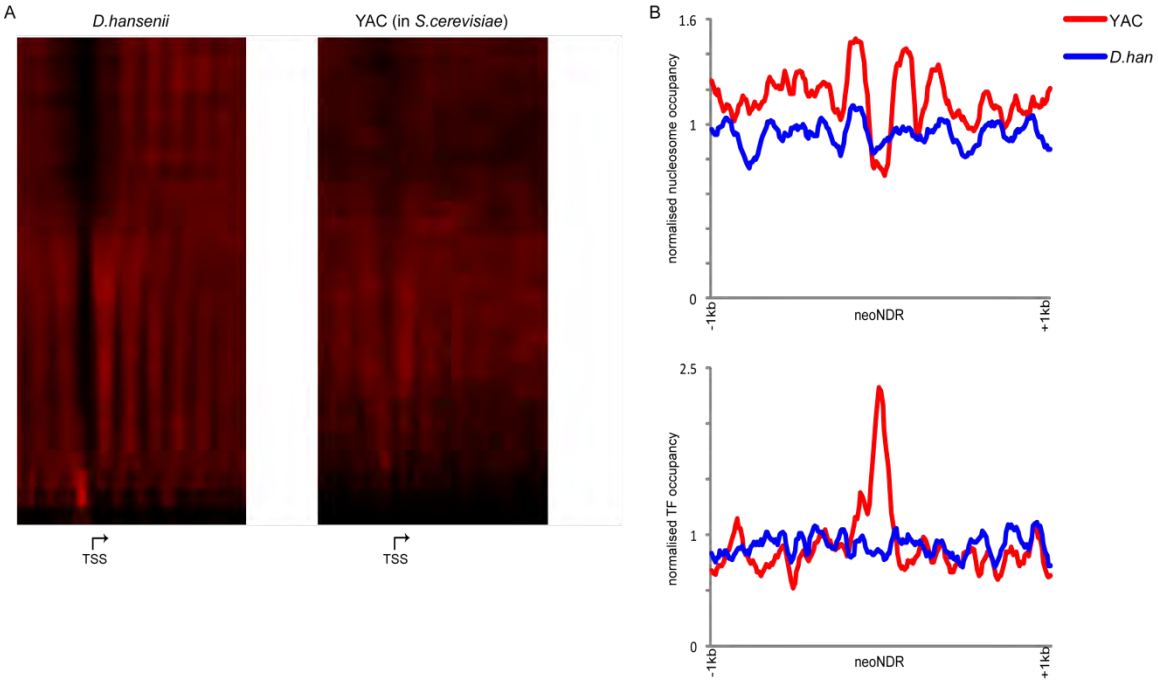


Figure IV. 2: Transcription-factor binding is associated with NDR formation.

(A) TSS aligned averages of all *D.hansenii* genes present on the YACs in their native context (left panel) and in the *S.cerevisiae* context (right panel), scaling from 20 to 300bp show loss of TF protection, along with previously documented NDR fill-in.
(B) Average of nucleosome-sized fragments surrounding *D.hansenii* YAC neoNDRs in the native (blue) and YAC (red) contexts (upper panel) and for 20-80bp footprints (bottom panel) show gain of TF footprints at previously defined YAC neoNDRs.

In order to determine what transcription factors might be involved in the differential footprinting seen on *D.hansenii* YACs, we searched for motif enrichment in transcription factor sized (20-80bp) footprints. DREME (Machanick and Bailey, 2011) analysis revealed that the Cbf1 motif was much more highly enriched (E-value of 7.4e-024) in TF footprints in *D.hansenii* than in *S.cerevisiae*. Cbf1 has previously been predicted to be a GRF associated with nucleosome

depletion in *D.hansenii* but not in *S.cerevisiae* (Tsankov et al., 2010). Interestingly, aligning TF footprints on the Cbf1 motifs in *S.cerevisiae* shows that these sequences are in fact bound in *S.cerevisiae* (Fig3B). However, the nucleosome profiles around Cbf1 motifs in *S.cerevisiae* show less nucleosome depletion and the less well-defined array than similar alignments in *D.hansenii* (Figure IV.3B). In order to determine if this was a gain-of-function in the *D.hansenii* Cbf1, we constructed strains that replaced the endogenous copy of *CBF1* in *S.cerevisiae* for that of *D.hansenii*. The deletion of *CBF1* in *S.cerevisiae* does lead to an increase in nucleosome occupancy at its binding sites in promoters (Figure IV.3C), consistent with a role for TF binding in promoter nucleosome depletion. Mapping of nucleosomes and TFs in the “Cbf1 swap” strain indicates that the *D.hansenii* copy of Cbf1 neither induces greater nucleosome depletion nor generates more well-positioned arrays at Cbf1-bound promoters in *S.cerevisiae*. This may indicate that Cbf1’s role as a GRF in *D.hansenii* is dependent upon some *D.hansenii*-specific association with another factor (or factors), or this may be due to the involvement of other *S.cerevisiae* TFs that help to induce the nucleosomal arrays around the binding site.

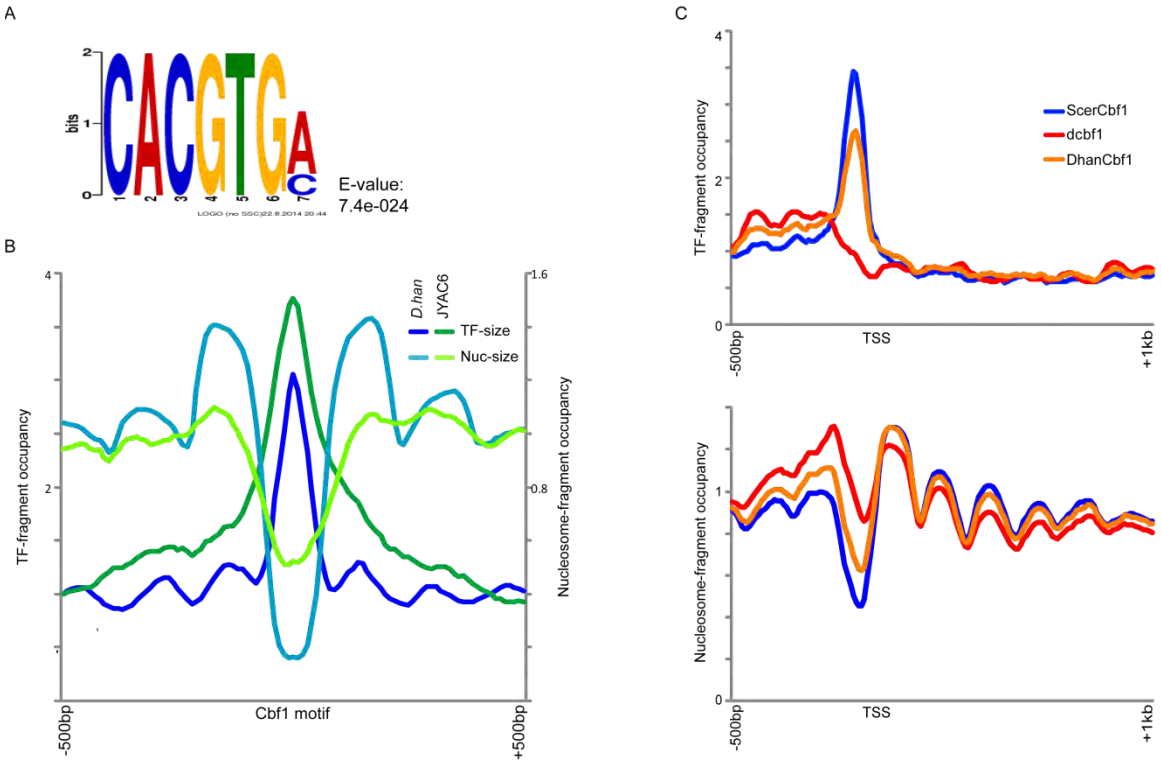


Figure IV. 3: Cbf1 acts as a GRF in *D.hansenii*, but does not possess inherent nucleosome ordering properties.

(A) DREME analysis of *D.hansenii* TF footprints over *S.cerevisiae* shows Cbf1 motif enrichment.

(B) Average alignments surrounding Cbf1 motifs present in YAC-born *D.hansenii* sequence show binding of the motif in *S.cerevisiae* as well as *D.hansenii*, but weaker nucleosome depletion and array formation on and around the motif.

(C) Average alignments of TF-sized (upper panel) and nucleosome-sized (bottom panel) fragments surrounding Cbf1-bound motifs for wild type, Δ*cbf1*, and *Dhan-CBF1* strains show binding of *D.hansenii CBF1* to these sites, but no increased nucleosome depletion or stronger array formation.

Discussion

Nucleosome depletion is formed by both *cis* and *trans* factors. Therefore nucleosome depletion can result from anti-nucleosomal sequence hindering the wrapping of DNA into a nucleosome, binding of sequence-specific factors abrogating nucleosome binding, or from the activity of chromatin remodelers. In a previous study (Hughes et al., 2012), we observe the loss and gain of NDRs on *D.hansenii* sequence depending on the environmental context, indicating that these NDRs are not *cis* regulated and thus excluding the former cause. When we systematically map TF and nucleosome footprints on *D.hansenii* and *S.cerevisiae* sequence, we see the enrichment of TF footprints in nucleosome depleted regions. This is consistent with a role for TFs in competing with nucleosomes and helping to generated NDRs in *trans*. As we would expect to follow from this, the NDRs that gain occupancy on *D.hansenii* sequence in an *S.cerevisiae* cell also lose binding of TFs, while the sequences that lose nucleosome occupancy in the *S.cerevisiae* environment gain footprinting of TFs here. This is consistent with our previous supposition that the generation neoNDRs in the YACs is due to the binding of a transcription factor that competes with the histone proteins. However, we cannot rule out the possibility that another factor disrupts the nucleosome in a species-specific context, and this allows for binding of TFs.

Previous nucleosome mapping across the yeast phylogeny has implicated general regulatory factors (GRFs) in generating nucleosome depletion and has further predicted the evolution of both general regulatory factors themselves as well as the frequency of their use in NDR generation (Tsankov et al., 2010). Consistent with the prediction of Cbf1 as a GRF in *D.hansenii* based on the enrichment of its motif in this species, we saw a greater enrichment of the Cbf1 motif in *D.hansenii* TF footprint than in *S.cerevisiae*'s. Interestingly, we saw that there were still TF footprints on the Cbf1 motif in *S.cerevisiae*, however these did not associate with as strong a nucleosome depletion or as well positioned a nucleosomal array as seen in *D.hansenii*. This suggests that the activity of Cbf1 has evolved differently in these two species. In order to see if NDR and nucleosomal array generation was a gain-of-function inherent to *D.hansenii* Cbf1, we generated an *S.cerevisiae cbf1* deletion mutant and a strain that was complemented with *D.hansenii* Cbf1. As expected for a role of TFs in excluding nucleosomes, we did see filling in of NDRs, where Cbf1 footprinting was lost in the deletion mutant. The *S.cerevisiae* strain with *D.hansenii*'s *CBF1* does not appear to generate any greater nucleosome depletion nor affect the positioning of nucleosomal arrays at Cbf1-bound promoters. This argues that the Cbf1 protein itself does not lead to the GRF activity in *D.hansenii*, but may be a result of recruitment of a remodeler (or other factor) in *D.hansenii*. It is possible that remodeler recruitment could represent one mechanism by which GRFs can distinguish themselves from run-of-the-mill TFs.

Materials and Methods

Yeast strains:

The strains bearing *D.hansenii* yeast artificial chromosomes were those used in ChapterIII and were generated by Yi Jin in the Struhl lab. JYAC6 contains a portion of chromosome C spanning 1165392-1280355 base pairs on a *URA3/TRP5* linearized vector, while JYAC7 inserted sequence houses 1148162-1364529 from chromosome D. These strains were compared to the laboratory *S.cerevisiae* wild type strain, BY4741, and *D.hansenii* strain NCYC 2572. Deletion of *CBF1* from diploid *S.cerevisiae* was performed via yeast homologous recombination with a *URA3* PCR product, tailed with 40bp of homologous sequence from either side of the *CBF1* coding region. Single colonies were selected for the deletion on URA- media and confirmed via colony PCR, before tetrad dissection and similar selection and confirmation. Replacing this gene with the *D.hansenii* copy was achieved also by yeast homologous recombination by amplifying the *D.hansenii* copy from genomic DNA with tailed primers that had 40bp of homology to the same *S.cerevisiae* sequence surrounding the *CBF1* coding region. Single colonies were selected via growth on MET- media (as the strains were genotypically methionine prototrophs, but *cbf1* deletion generates a phenotypically methionine auxotroph).

Primer	Sequence	Purpose
SCbf1::URA3 5’	AATACGGTTTTCTACACTTTTATTAACGATGAA CTCTCTGGGATTCGGTAATCTCCGA	Cbf1 deletion
SCbf1::URA3 3’	TTTAACTCTCAAGCCTCATGTGGATTATCGCT CCTAGTGCGGGTAATAACTGATATAATT	Cbf1 deletion
SCbf1::DCbf1 5’	CTTAAAATATAATACGGTTTTCTACACTTTTATT AACGATGTCGAAAAGATCATC	Cbf1 swap
SCbf1::DCbf1 3’	GAGACTCGAAATACATTTAGCTATCTATTTTTA ACTCTCATTTCATATTCTTTTCGTCC	Cbf1 swap
Cbf1 55up 5’	CAAGTACCAACATCAAGTGC	PCR confirm
Cbf1 41 down 3’	CAGATACATAGGGAGACTCG	PCR confirm
Cbf1-Dhan in 3’1	ATCGCAATCTCGTCCTTC	PCR confirm (swap)
Cbf1-Dhan in 5’2	CGAACGACAAGTTGAAGC	PCR confirm (swap)
URA3 5’3	GGGTGTATACAGAATAGCAGAATGGGCAGA	PCR confirm (deletion)
URA3 R1	TTGGCGGATAATGCCTTTAGCGGCTTAACT	PCR confirm (deletion)

Table IV. 1: Primers used for Cbf1 strain construction

Cbf1 deletion primers were used in a URA3 PCR reaction from yeast plasmid pRS416 and the PCR product was used in a high efficiency transformation to replace *CBF1* coding region with *URA3*. Cbf1 swap primers amplified *D.hansenii* genomic DNA, and the PCR product was transformed into the deletion mutant to incorporate the *D.hansenii CBF1* coding sequence into the *S.cerevisiae CBF1* genomic location. Cbf1 55up 5’ was used in tandem with Cbf1-Dhan in 3’1 and URA3R1 to confirm Cbf1 swap or deletion, respectively, while Cbf1 41down 3’ was used with Cbf1-Dhan in 5’2 and URA3 5’3 to PCR the other side of the swap or deletion, respectively.

Yeast Growth:

Yeast were grown in a version of compromise media (0.2% Synthetic Complete – Trp –Ura, 1% bacto-peptone, 0.5% yeast extract, 2% dextrose) which was used across the yeast species (Tsankov et al., 2010); the media was modified to have reduced tryptophan and uracil, as the yeast artificial chromosomes bore *URA3*

and *TRP5* selectable markers. Prior to liquid growth, yeast artificial chromosome strains were grown on AHC plates, which was more stringent for YAC retention. Yeast cultures were grown at 30°C shaking at 220rpm in an Innova 44 shaker.

Micrococcal Nuclease Digested DNA Isolation:

Yeast cultures (200mL) were fixed with 1.85% formaldehyde at 30°C, shaking, for 30 minutes. Yeast cultures were pelleted and washed. Cell pellets were resusupended in 1.2mL of cell breaking buffer (10mM Tris, pH 7.4, 20% glycerol) with Sigma protease inhibitors and split into 2 screw cap tubes with approximately 0.6mL of 0.5mm zirconia/silica beads (Biospec Products 11079105z). The cell wall was broken in a Mini-Beadbeater-96 (Biospec Products) in a cold magnetic rack by shaking twice for 2 minutes on the bead beater and an additional 1 minute and 17 seconds on the bead beater, each of these was separated by 1 minute and 15 seconds on ice. The cells were checked under a microscope to ensure that most were broken; the lysed cells were separated from the beads by spinning them through a puncture in the bottom of the tubes into 5mL tubes. The cells were spun at maximum speed in a 4°C eppendorf centrifuge and the supernatant was discarded. The cells were resuspended up to about 2.4mL with NP buffer (50mM sodium chloride, 10mM Tris pH7.4, 5mM magnesium chloride, 1mM calcium chloride, 0.5mM spermidine, 1μL/mL β-mercaptoethanol, and 0.01% NP-40) and 0.6mL of cells was aliquoted to four tubes with micrococcal nuclease aliquoted in the lid. The cells were mixed

with the enzyme and incubated at 37°C for 20 minutes, after which 150µL of STOP buffer (50% sodium dodecylsulfate, 0.05M ethylenediaminetetraacetic acid) and 5µL 20mg/mL proteinase K and the digestions were incubated at 65°C overnight to reverse crosslinks and remove proteins. Mononucleosomal DNA was purified by a PCI extraction and precipitation with 0.3M sodium acetate in 2-propanol and then dissolved in 60µL of 1X NEB buffer 2. RNA was removed with 2µL of 20mg/mL RNase solution (Sigma) at 37°C for 1 hour. Digestion ladders were assessed by running 5µL on a 2% agarose gel. Digestions were chosen to have mostly mononucleosomal-sized DNA with a little dinucleosomal and a hint of trinucleosomal DNA. Half (25µL) of the appropriate digestion was treated with calf intestinal phosphatase (0.75µL) at 37°C for 45 minutes. This was cleaned up with a PCI extraction and ethanol precipitated.

Deep Sequencing Library Preparation:

Estimated 1000ng of digested DNA was end-cleaned with END-it (Epicentre) in a 40µL reaction. The reaction was cleaned-up via PCI extraction and ethanol precipitation for at least 1 hour and was then A-tailed with Klenow exo-minus DNA Polymerase (Epicentre) in a 25µL reaction. This reaction was similarly cleaned up. A 15µL ligation reaction (Fast Link, Epicentre) was performed at room temperature for 1 hour and the reaction mixture was made up to 25µL and the ligation was continued at 16°C overnight. The ligation reaction was cleaned up with 1.8X Agencourt AMPure XP beads (Beckman Coulter), washed twice

with 70% ethanol, and eluted with 39μL water. Two-thirds of the ligated DNA was used for two 25μL Pfx PCR reaction, using 10 and 12 cycles of the PCR extensions. A portion of the PCR reactions (4μL) was run on a gel to check the presence and size of the product. A final PCR with the remaining one-third of the ligated material was repeated with the best cycling conditions for each sample. PCR reactions were pooled and mixed with 1.8X AMPure beads, washed on the beads two times with 70% ethanol, and eluted in 20μL water. One tenth of the purified library was used to quality check the library by StrataCloning, which requires an initial A-tailing reaction with Taq DNA Polymerase. The libraries were submitted to the UMass deep sequencing core to be mixed at 1:1 molar ratios and sequenced on a paired-end Hi-Seq lane.

Data Analysis:

Raw, fastq files were separated by barcode using Novobarcode, and the ends were trimmed of adaptor read-through sequence with a homemade perl script. Reads were aligned to the sacCer3 genome (or to the *D.hansenii* chromosome sequences CR382133-9, sequenced 2008/09/10, downloaded from genolevures.org) using bowtie2 with the –dovetail option. For the yeast artificial chromosomes strains, reads were aligned to a chimeric genome containing the whole sacCer3 genome along with the two *D.hansenii* chromosomes. The aligned reads were filtered to keep uniquely mapping concordant reads and separated into size classes, representing transcription factors (~20-80bp) and

nucleosomes (~120-160), as well as separating into 10bp binned size classes 20-300bp (for V-plots). In order to generate transcription start site alignments, the TF and nucleosome size class reads were extended across the fragment size (using either end of the paired-end read data) and each base pair in all fragments was counted and totaled, using a homemade perl script. The genome-wide average base pair count was normalized to one, and the normalized reads were used to generated TSS alignments 500bp upstream to 1000bp downstream of the edge of the +1 nucleosome as defined in Tsankov et al. (Tsankov et al., 2010). Similar alignments were generated surrounding the previously defined "neoNDRs" from Hughes et al. (Hughes et al., 2012) and for Cbf1 motifs. To make V-plots, the 10bp binned fragments were similarly counted and gene alignments were made and averaged across genes for each size class and viewed in Java TreeView.

In order to determine motifs in the TF footprints, the bowtie2 aligned SAM files were mapped to the relevant genome or chimeric genome on SeqMonk, which was used to generate probes that represented locations of high protection (the top 10% of probe values for 20-80bp fragments). These probes were further filtered to remove ones that were highly protected in nucleosome-sized fragments. The sequence at these locations was extracted from the chromosome sequence and was used in DREME (Machanick and Bailey, 2011) to look for enrichment in *D.hansenii* over *S.cerevisiae* footprints.

CHAPTER V: Chd1p is a *trans*-acting factor that sets nucleosome spacing

Introduction

Nucleosomes wrap ~147bp of DNA in almost 2 turns around histone proteins and form the basic repeating unit of chromatin. Chromatin affects the accessibility of the underlying DNA sequence, and hence, where nucleosomes are positioned is important to the regulation of the genome. This has been particularly notable at yeast genes, where nucleosomes are depleted over the promoter and form a well-positioned array over the gene body, almost universally. While sequence has been shown to play a role in nucleosome depletion at promoters (Kaplan et al., 2009), *trans*-acting factors play a larger role in nucleosome positioning across the genome (Zhang et al., 2011).

ATP-dependent chromatin remodelers contain a SNF2 ATPase domain that provides the energy to disrupt histone-DNA contacts. There are four major classes of these remodelers: SWI/SNF, INO80, ISW, and, CHD. Along with the ATPase motor, these remodelers all carry accessory domains, and interact with additional subunits, that contribute to different outcomes of remodeling events *in vitro* (Clapier and Cairns, 2009). SWI/SNF remodelers are often associated with histone eviction or movement in *trans* to another DNA molecule, while ISW and CHD remodelers result in sliding in *cis* along DNA. SWR/INO80 play key roles in

altering the subunit composition of nucleosomes, as they exchange the histone variants H2A and H2A.Z in the nucleosome.

In *vitro* assays have shown directionality of remodeler activities: Isw1a, Isw2, and Chd1 all lead to centering of nucleosomes on a DNA fragment (Stockdale et al., 2006), while Isw1b and human SWI/SNF move nucleosomes to the ends of DNA (Bouazoune et al., 2009). The composition of the accessory domains/subunits of these remodeling complexes is assumed to play a role in the different outcomes of the remodeling reactions. Accessory domains are involved in recognizing histone modifications or variants as well as DNA, and regulate the activity of remodelers (Clapier and Cairns, 2009). For instance, Isw2, Isw1a, and Chd1, all preferentially bind nucleosomes with extranucleosomal DNA, leading to the hypothesis that this preference may be integral to their centering activity (Hota et al., 2013). In fact, replacing the DNA binding domain of Chd1 with a domain targeting other nucleosomes resulted in an activity that shifts nucleosomes to the edge of a DNA molecule (Patel et al., 2013).

Chd1 is composed of N-terminal tandem chromodomains, an ATPase domain, and C-terminal SANT/SLIDE DNA binding domain connected to the ATPase domain by a linker. Both the chromodomains and the DNA binding domain have been crystallized with the ATPase domain. In the former, the chromodomains were seen to fold against the two lobes of the motor, seemingly holding them in a configuration that is not consistent with ATP hydrolysis, and

obscuring a DNA binding surface on the ATPase domain (Hauk et al., 2010). These structural observations are consistent with the negative regulation of Chd1 activity by the N-terminus seen in *in vitro* assays; this regulation is overcome in the presence of histones. On the other hand, the DNA binding domains appear to positively regulate the activity of Chd1 *in vitro*. The DNA binding activity of Chd1 may also be involved in directing sliding of nucleosomes, as artificial fusion of Chd1 with a sequence-specific DNA-binding domain slides nucleosomes towards the DNA binding site (McKnight et al., 2011). The linker connecting the DNA binding domains to the ATPase motor has been shown to be required for coupling ATP hydrolysis to nucleosomal movement (Patel et al., 2011).

Deletion of Chd1 in *Saccharyomyces cerevisiae* leads to loss of nucleosome positioning at the 3' end of genes. This becomes even more extreme when this deletion is coupled with deletions of Isw1 and Isw2, as this triple deletion mutant exhibits complete loss of nucleosomal phasing after the +2 nucleosome (Gkikopoulos et al., 2011). Consistent with these observations from budding yeast, the two *Schizosaccharyomyces pombe* orthologs of Chd1 also play central roles in nucleosomal array formation (Pointner et al., 2012). Furthermore, Chd1 is involved in suppressing aberrant histone turnover in the gene body of long genes (Radman-Livaja et al., 2012; Smolle et al., 2012). Despite its stronger role in gene body nucleosome positioning, there is conflicting evidence as to the localization of Chd1 to genes. While crosslinked Chd1 has been seen across the coding region of some genes (Simic et al., 2003), native

ChIP shows it to be primarily enriched at the promoter region (Zentner et al., 2013). These seemingly incompatible results may be explained by two observations regarding Chd1 binding. Chd1 genetically interacts with yFACT and elongation factors (Biswas et al., 2007; Simic et al., 2003), and these factors have been proposed to help localize Chd1 to coding regions. Chd1 association with the gene body may be transient, given its preference for binding nucleosomes with a longer stretch of extra-nucleosomal DNA than the average yeast linker DNA. Despite the knowledge of Chd1 and its regulatory domains' *in vitro* activity, how these influence *in vivo* chromatin organization remains unclear.

Previously we have used a yeast artificial chromosome (YAC) system to identify those chromatin characteristics that are influenced by *trans* regulation (Hughes et al., 2012), based on aspects of chromatin that differ when the same DNA sequence is carried in two different environments (*K. lactis* vs. *S.cerevisiae*). From this study, we established that the majority of nucleosomes are positioned by trans-acting factors, as spacing of nucleosomes was environmentally sensitive (internucleosome distance on *K.lactis* sequence adopted the tighter spacing of the *S.cerevisiae* genome). The existence of a "molecular ruler" whose measurement differs between the species allows for a complementation assay to determine what factor determines the measurement between nucleosomes. We therefore generated *S.cerevisiae* "factor swap" strains, in which deletion of an endogenous gene was complemented with the *K.lactis* ortholog, and performed MNase-Seq to map nucleosomes in these

strains. We confirmed that deletion of Chd1 lead to loss of 3' nucleosome positioning, and found that *K.lactis* Chd1 was able to complement this loss of positioning and additionally generated nucleosomal arrays with increased spacing. Generation of chimeric Chd1 genes revealed that the Chd1 N-terminus is responsible for much of this increased linker length.

Results

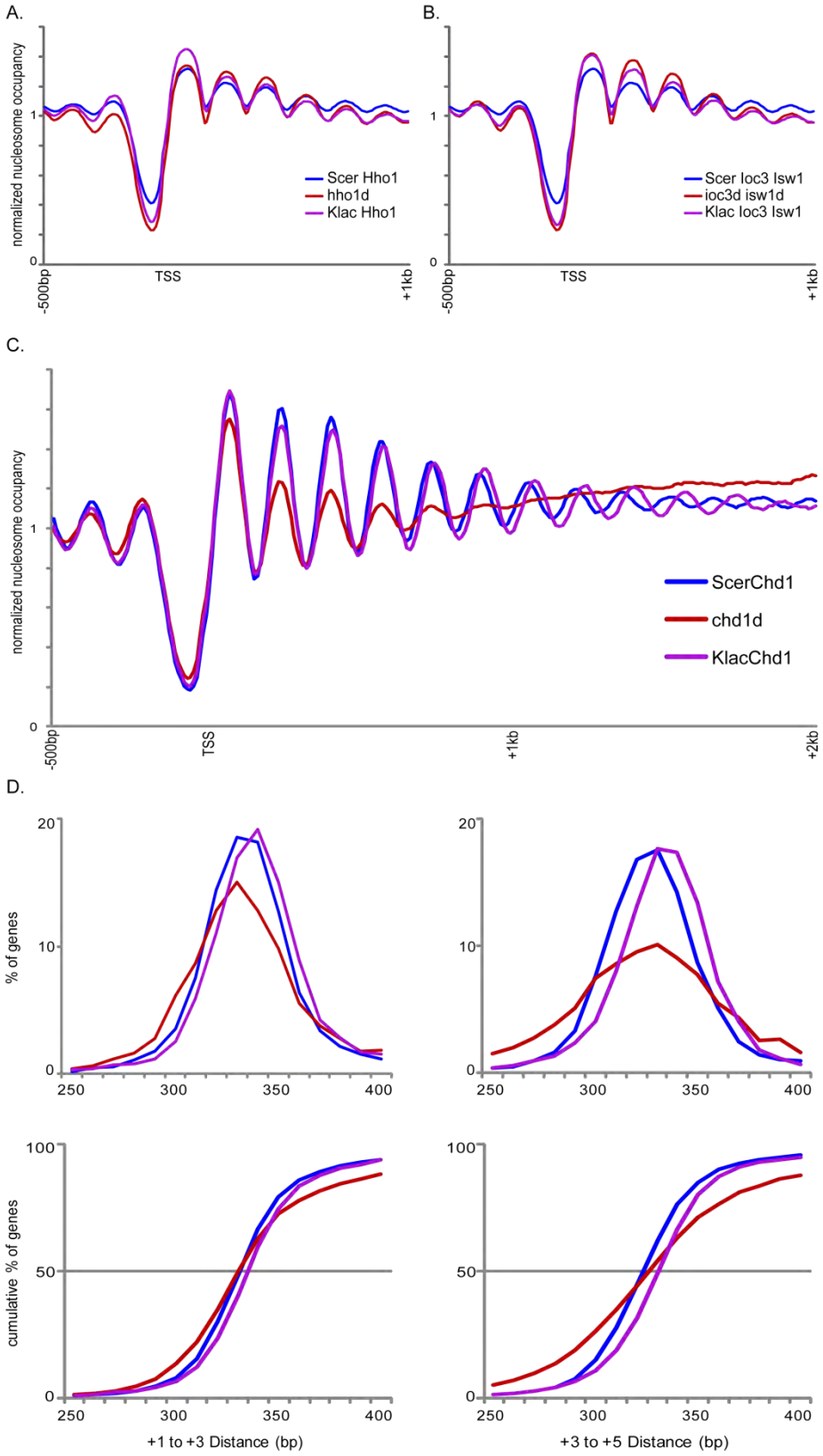
In order to test chromatin remodelers for internucleosomal spacing function, we deleted candidate factors from *Saccharomyces cerevisiae* and complemented them with the gene from *Kluyveromyces lactis*, a species that exhibits ~15bp greater distance between coding region nucleosomes than does *S.cerevisiae* (Heus et al., 1993; Tsankov et al., 2010). By mapping nucleosomes in the deletion and complementation strains, we can potentially identify proteins responsible for the differential measurement between nucleosomes in these species. Several observations in the literature suggested that histone H1 might be responsible for an organism's average linker length (Heus et al., 1993). Curiously, recent physical models of fungal nucleosome spacing were able to successfully model internucleosomal spacing and apparent nucleosome width for 11 out of 12 species, with *K.lactis* being the lone failure (Möbius et al., 2013) – the observation was hypothesized to result from high levels of H1 interfering with "breathing" of DNA on nucleosomes in this species. However, deletion of the linker histone, Hho1, from *S.cerevisiae* and its replacement with the *K.lactis* copy

did not alter the nucleosome phasing pattern across the yeast genes (Figure V.1A).

We next considered the hypothesis that the Isw1 remodeler might be responsible for the difference in nucleosome spacing between these species. The crystal structure of the DNA binding portions of Isw1a suggested that this complex could measure the distance between nucleosomes due to its contact with DNA at both the entry and exit points of the nucleosome (Yamada et al., 2011); this is consistent with its spacing of nucleosome arrays *in vitro* (Vary et al., 2003). However, neither the deletion nor complementation of Ioc3 and Isw1 (Isw1a complex) altered nucleosome spacing in *S.cerevisiae* (Fig1B). Other than a slight increase in 5' nucleosome occupancy, the deletion of the Isw1a complex does not alter average genic nucleosome organization.

Figure V. 1: K.lactis Chd1 generates increased internucleosome spacing throughout coding regions

A) Average TSS aligned nucleosome profiles show that neither Hho1 deletion nor swap affects nucleosome spacing on *S.cerevisiae* DNA.
B) As in (A), but for Isw1a.
C) Average TSS aligned nucleosome profiles show Chd1 deletion causes loss of 3' nucleosome positioning. *K.lactis* Chd1 generates increased internucleosome spacing.
D) Histograms (upper panels) and cumulative distribution plots (lower panels) show that +1 to +3 nucleosome distance (left panels) is increased by *K.lactis* Chd1, but to a lesser extent than for +3 to +5 nucleosomes (right panels).


A.
normalized nucleosome occupancy
1
0
Scer Hho1
hho1d
Klac Hho1
-500bp
TSS
+1kb
B.
1
0
Scer Ioc3 Isw1
ioc3d isw1d
Klac Ioc3 Isw1
-500bp
TSS
+1kb
C.
normalized nucleosome occupancy
1
0
ScerChd1
chd1d
KlacChd1
-500bp
TSS
+1kb
+2kb
D.
% of genes
20
10
0
250
300
350
400
cumulative % of genes
100
50
0
+1 to +3 Distance (bp)
+3 to +5 Distance (bp)

Deletion of Isw1 along with Chd1 disrupts nucleosome positioning downstream of the +2 nucleosome (Gkikopoulos et al., 2011). Additionally, Chd1 has nucleosome spacing activity *in vitro* (Stockdale et al., 2006), so we chose to test the hypothesis that Chd1 plays a role in linker measurement in vivo. We confirmed published results showing that deletion of Chd1 results in disorganized 3' nucleosome positioning (Figure V.1C) (Gkikopoulos et al., 2011). Interestingly, complementation with *K.lactis* Chd1 causes an increase in nucleosome spacing at the 3' end of genes (Figure V.1C). By calling nucleosome peaks and determining the distance between adjacent nucleosomes, we see that nucleosomes at the 5' end of genes are spaced farther apart in the presence of *K.lactis* Chd1, although to a lesser extent than nucleosomes farther into the gene body (Figure V.1D). Chd1 thus appears to be involved in the measurement of internucleosome distances, and evolutionary divergence between the *S.cerevisiae* and *K.lactis* copies results in altered measurement of linker length. However, there are other factors present that influence linker length, as the increase in nucleosome spacing seen with the introduction of *K.lactis* Chd1 does not fully account for the longer linker lengths seen in the *K.lactis* genome.

Chd1 affects spacing of gene body nucleosomes, and it has previously been observed to genetically and physically interact with yFACT as well as the Paf1 complex (Biswas et al., 2007; Simic et al., 2003). It has been hypothesized that Chd1 exerts its effects on gene body nucleosomes by binding to the longer extranucleosomal DNA present as elongating RNA polymerase exposes

nucleosomal DNA (Zentner et al., 2013). If Chd1's function is targeted to gene body nucleosomes via RNA polymerase elongation, then the increased linker length in the presence of *K.lactis* Chd1 should be stronger at more highly transcribed genes. Separating genes into quartiles based on RNA Polymerase II occupancy in their coding regions revealed that the most highly-expressed genes exhibit a greater increase in nucleosome spacing in the presence of *K.lactis* Chd1 than the lowest expressed genes (Figure V.2A). Interestingly, the effects of *K.lactis* Chd1 do not appear to differ between those genes with the highest and lowest occupancy of Chd1 from native ChIP (data not shown). These results do not exclude the possibility that Chd1 primarily functions during transcription to help set the ordered spacing of nucleosomes in the array, but argues that either the continued presence of Chd1 is not required for its activity in the gene body or active Chd1 does not stably interact with gene bodies without crosslinking.

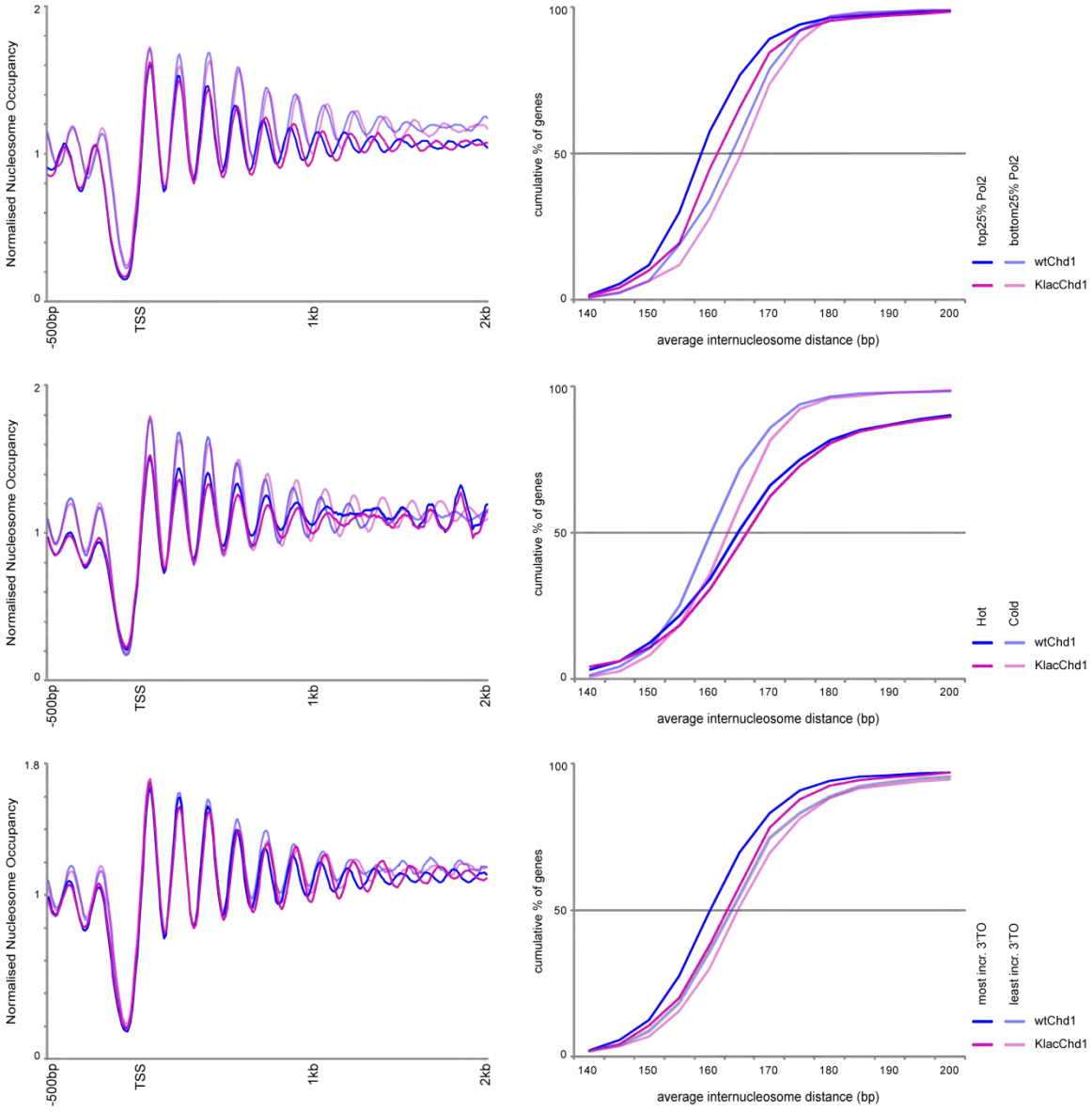


Figure V. 2: Chd1 spacing activity is greatest at the most highly-expressed genes.

A) Average TSS alignment profiles (left) and cumulative distribution plots (right) of internucleosome distances show that there is a greater increase from wild type *S.cerevisiae* to *K.lactis* swap Chd1 copy at genes with the greatest Pol2.

B) As in (A), but showing that genes with the coldest gene body nucleosome have a slightly greater increase in spacing with *K.lactis* Chd1.

C) As in (A), but showing the genes whose 3' turnover is most affected by Chd1 deletion are also more affected by Chd1 copy.

We next compared data from our Chd1 swap strains with published datasets which assayed the effect of Chd1 on replication-independent histone replacement (Radman-Livaja et al., 2012; Smolle et al., 2012). Both of these studies found that loss of Chd1 lead to an increase in histone turnover at the 3' ends of long genes, suggesting that this remodeler acts either to stabilize nucleosomes during transcription or to aid in their reassembly after transcription. Interestingly, those genes which show the greatest increase in 3' turnover in *chd1* mutants also exhibit a greater increase in spacing in strains carrying the *K.lactis* Chd1 (Figure V.2C) – these genes thus are the most responsive to both primary activities of Chd1 *in vivo*. However, *K.lactis* Chd1 had no greater effect on increasing nucleosome spacing in genes with the highest wild-type gene body turnover than in genes with the lowest turnover (Fig2B), indicating that turnover is unlikely to be the primary determinant of Chd1 access and activity.

Having identified a pair of Chd1 orthologs that direct different internucleosome spacing, we sought to generate chimeric Chd1 proteins in order to identify the "molecular ruler" responsible for linker length. ATP-dependent chromatin remodelers bear a similar ATPase motor, but have distinct accessory domains/subunits that recognize DNA and histone modifications and variants. These domains regulate the localization and activity of the remodelers. In Chd1, these accessory domains include tandem chromodomains at the N-terminus, and DNA binding domains at the C-terminus that flank the central helicase. The chromodomains have been shown to negatively regulate the ATPase activity of

Chd1 in the absence of histones, while the DNA binding domains positively regulate its activity (Hauk et al., 2010). It had been postulated that the DNA binding domain served to anchor the remodeler and generate torsion in a power stroke to move nucleosomes; however the rigid coupling of the DNA binding domain to the ATPase was shown to be unnecessary for nucleosome remodeling, arguing against the power stroke model for Chd1 and *Drosophila* ISW (Ludwigsen et al., 2013; Nodelman and Bowman, 2013). Interestingly, a longer linker does allow for a site-specific Chd1 to remodel nucleosomes farther away from its binding site (Nodelman and Bowman, 2013). In order to test which domain(s) of Chd1 are responsible for its measurement activity, we generated chimeras between *K.lactis* and *S.cerevisiae* Chd1 (Fig3A), and tested these chimeric proteins for longer measurement in the *S.cerevisiae* genome.

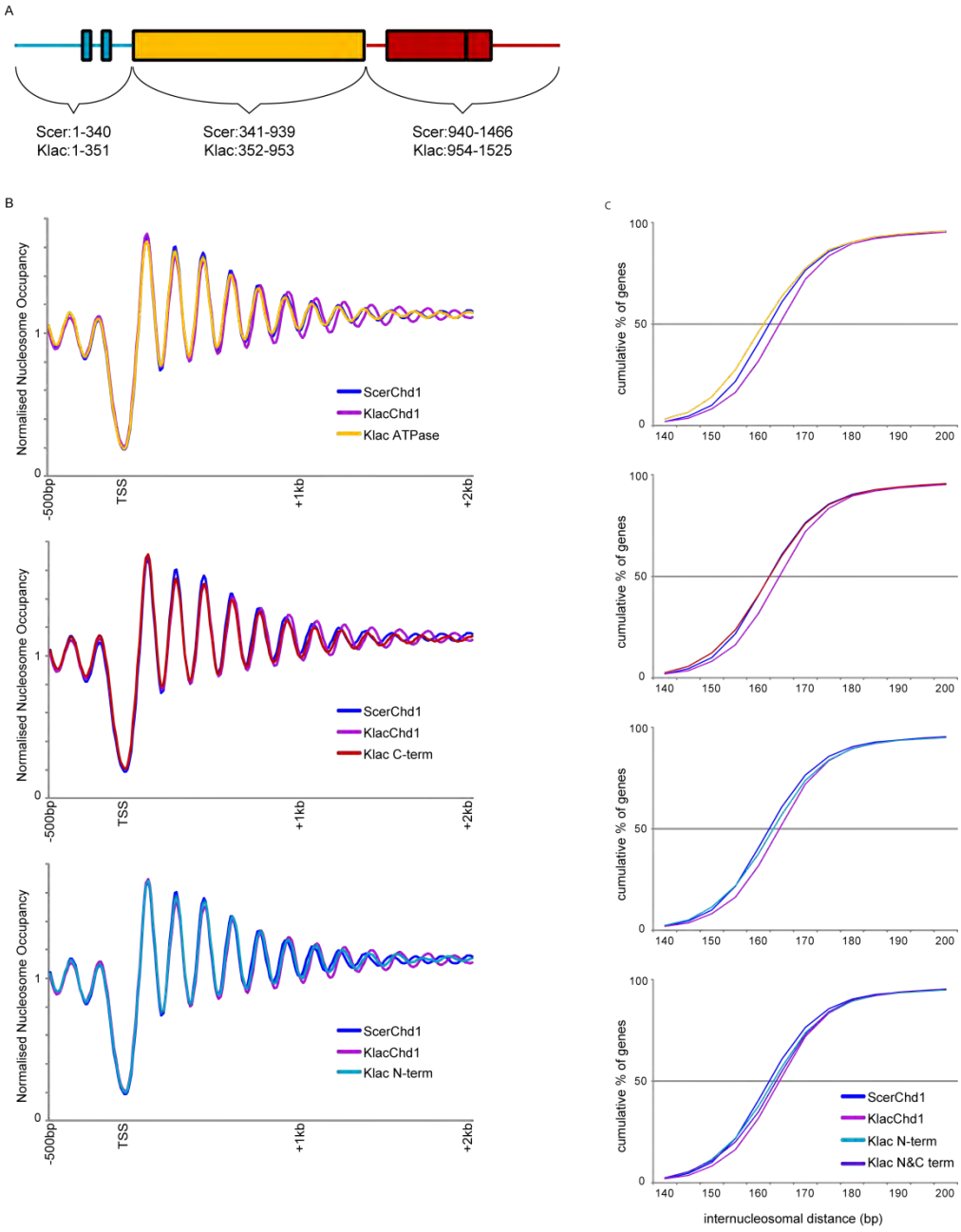


Figure V. 3: The N-terminus of Chd1 bears the majority of the differential measurement in the species.

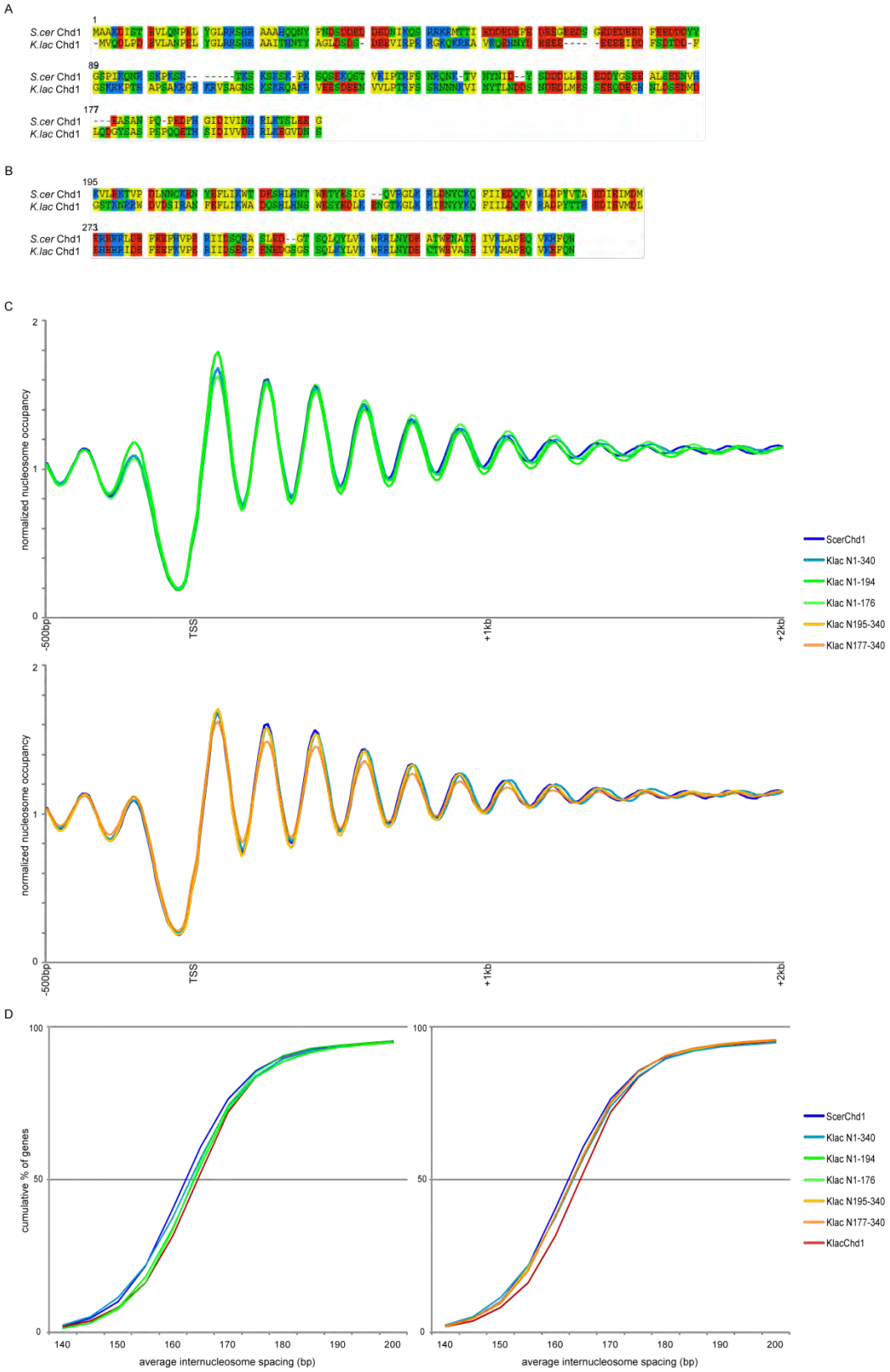
A) Schematic of Chd1 chimerae divisions.
B) TSS alignments of ATPase (left), C-terminal (middle), and N-terminal (right) chimerae show that the ATPase domain does not contribute to differential nucleosome spacing, while the C- and N-termini contribute to spacing differences.
C) Cumulative distribution plots support the differential spacing seen for TSS alignments (B).

As expected, the highly conserved ATPase domains did not generate any increase in linker measurement (Figure V.3B). Replacing the C-terminus from *S.cerevisiae* Chd1 with that from *K.lactis* caused a very slight increase in nucleosome spacing (Fig3B). The greatest effect of a single domain swap was observed in chimeric proteins carrying the N-terminus of Chd1 from *K. lactis* (Figure V.3B), although even in this mutant we did not recover the full effect of the full length protein swap. The N-terminal chromodomains have been shown to regulate the activity of Chd1 by interfering with the association of the ATPase lobes; it has been proposed that histone binding induces a conformational change that allows the helicase domain to bind DNA and hydrolyze ATP (Hauk et al., 2010). While chromodomain2 is well conserved between *S.cerevisiae* and *K.lactis* chromodomain1 is slightly more divergent (Figure V.4B). Interestingly, the N-terminal portion of Chd1 (residues 1-118) is typically truncated in proteins used for biochemical analysis, so little is known about their involvement in Chd1 structure and function. In order to test the involvement of the chromodomains and N-terminus, we generated chimerae for the N terminus. Surprisingly, we saw that swapping N-terminal portion of *S.cerevisiae* Chd1, comprising amino acids 1-195, for the *K.lactis* sequence led to a greater increase in average genic nucleosome spacing than did the *K.lactis* chromodomains (Figure V.4C&D). While the increased measurement between nucleosomes in *K.lactis* appears to be concentrated in the N-terminal portion of Chd1, the spacing activity is distributed throughout the protein, as the full length *K.lactis* Chd1 has a greater

effect than any chimera on *S.cerevisiae* nucleosome spacing. The N-terminal portion of Chd1 could contribute to the measurement differences between the species by differentially regulating the ATPase activity in the presence of histones, or by altering the recruitment or abundance of the protein.

Figure V. 4: N-terminus but not chromodomains is required for the increased measurement seen in the N-terminal chimera.

A) Alignment of N-terminal 196 amino acids of *S.cerevisiae* and *K.lactis* Chd1.
B) Alignment of *S.cerevisiae* and *K.lactis* Chd1 chromodomains.
C) Average TSS alignment profiles show that swapping the amino acids N-terminal of (upper) the chromodomains, but not the chromodomains (lower), of *S.cerevisiae* Chd1 for *K.lactis* generate greater nucleosome spacing.
D) Cumulative distribution plots of nucleosome peak spacing agrees with the N-terminal amino acid-dependent increases seen in the TSS alignments (C).


A
S.cer Chd1
K.lac Chd1
B
S.cer Chd1
K.lac Chd1
C
normalized nucleosome occupancy
-500bp
TSS
+1kb
+2kb
ScerChd1
Klac N1-340
Klac N1-194
Klac N1-176
Klac N195-340
Klac N177-340
D
cumulative % of genes
average internucleosome spacing (bp)
KlacChd1

Discussion

Here, we present evidence that Chd1 is involved in spacing nucleosomes and in particular that it plays a role in the difference between internucleosome spacing in the yeast species, *S.cerevisiae* and *K.lactis*. This shows that a chromatin remodeler is involved in actively spacing nucleosomes. We see that Isw1a is not involved in the measurement differences between the species, but does not preclude this complex, which has *in vitro* spacing activity, from a conserved role in internucleosome spacing. In fact, clearly other factors play a role in the increased spacing seen in *K.lactis*, as Chd1 alone does not generate *K.lactis* spacing in the *S.cerevisae* background.

Our finding that *K.lactis*' Chd1 has a greater effect in increasing genic nucleosome spacing at the most highly expressed genes corresponds well with Chd1 association with transcribed genes. This could be due to interactions with elongation associated factors or from binding to transiently increased extranucleosomal DNA from transcription-associated histone turnover. *K.lactis* Chd1does not have an increased effect on nucleosome spacing for genes with higher turnover as would be expected if histone turnover were necessary for Chd1 activity, in fact genes with decreased turnover tend to exhibit greater *K.lactis* Chd1 induced spacing increases. Since Chd1 is involved in the repression of 3' turnover (Radman-Livaja et al., 2012; Smolle et al., 2012), it is possible that Chd1 itself, or other remodelers such as Isw1, might mask

transcription-induced turnover that allows for Chd1 binding to extranucleosomal DNA. However, the genes with the greatest turnover in the *chd1* deletion strain are no more affected than genes with the least amount of turnover. Our results do support the recruitment of Chd1 by transcription factors, such as Paf1 or yFACT, and do not rule out transcription-associated exposure of extra-nucleosomal DNA via partial unwrapping.

While the spacing activity of Chd1 is not completely localized to a single domain, the N-terminus of Chd1 appears to bear most of the increased measurement seen in *K.lactis*. This portion of the protein contains tandem chromodomains that have been observed to participate in negative regulation of Chd1's sliding activity in the absence of histones. However, the majority of the N-terminus' contribution to increased spacing appears to be N-terminal of the chromodomains, a region that has not been structurally characterized and is generally lacking in biochemical assays. The N-terminus of *K.lactis* Chd1 could potentially affect the folding of the protein and could lead to differential recognition of histone tails and thus alter regulation of ATPase activity by the chromodomains or its recruitment. Alternatively, the N-terminus of *K.lactis* Chd1 could affect the stability of the protein, and the differences between *K.lactis* and *S.cerevisiae* internucleosome spacing could result from different protein abundance.

Materials and Methods

Yeast strains:

Yeast strains were derived from the diploid S288C strain, BY4743, in which one copy of the HIS3 gene had been restored. Deletion of the coding region was performed via high efficiency transformation with a URA3 PCR product bearing homology to either side of the coding region. The deletion mutation was selected for on URA- media and was confirmed via colony PCR for single colonies. After the deletions were confirmed, cells were patched out onto YPD plates with high (5%) glucose for up to 16 hours at 30°C and were then sporulated at room temperature in sporulation media (1% potassium acetate, 0.1% yeast extract, 0.05% glucose) for several days. Once enough ascii were formed, as assessed by microscopy, they were tetrad dissected and replica plated to plates to confirm their genotype for the relevant deletion by growth on URA- media as well as LYS- and HIS- media, in order to compare strains with the same auxotrophies: all deletion strains were *leu2∆0*. For *K.lactis HHO1* and *ISW1IOC3* complemented strains, these genes were amplified from *K.lactis* genomic DNA with primers containing restriction enzyme sites; the resultant product was digested and ligated into the pRS415 yeast centromeric plasmid with LEU2 marker (HIS3 marker for *IOC3*). The deletion strains were transformed with either vector alone or the plasmid bearing the *K.lactis* gene. For Chd1, the *K.lactis* complementation strains were made via homologous recombination of the *K.lactis* PCR product

with 40bp of homology to either side of the *S.cerevisiae* gene in the haploid deletion strain. Transformants with this product were selected on 5-FOA media after an overnight outgrowth in YPD and integration was confirmed via colony PCR. Additionally, for a wild type comparison, an *S.cerevisiae* PCR product was reintegrated into the deletion strain and selected as for the *K.lactis* complementation.

Primer	Sequence	Purpose
Hho1-for	TATTTATGGGCACCTGATAATGCTTGGCAGCGAG GGAAGC*GATTCGGTAATCTCCGA*	Hho1 deletion
Hho1-rev	GTTTGATAGTATTGCTATCACCATTGACATTCTCG TTTGG*GGGTAATAACTGATATAATT*	Hho1 deletion
Isw1-for	AGCTATGCAAAAACCAGCTAGAGGTGGATGTAG AAATACC*GATTCGGTAATCTCCGA*	Isw1 deletion
Isw1-rev	TAGTATGATTATATATTTTTCTTCAGAAGCATGGT GTAGG*GGGTAATAACTGATATAATT*	Isw1 deletion
Ioc3::KANRF	GCAGCTCTTCCGCACAGCTTGCTGACCATGATCA TGATCG*TTCCCCGAAAAGTGCCACCTG*	Ioc3 deletion
Ioc3::KANRR	GCCTGTAAGGAGTTTCACAATCTTCACGTTCGTT GAAAGC*CGACAGCAGTATAGCGACCAG*	Ioc3 deletion
Chd1::URA3 5’	CCTTTTCTAATTTAATTCTCACTTATAATGGCAGC CAAGG*GATTCGGTAATCTCCGA*	Chd1 deletion
Chd1::URA3 3’	AAAATTGTTTCACTTCTTTTGAGACTCTGTTATCT TGTCG*GGGTAATAACTGATATAATT*	Chd1 deletion
Hho1_XhoIF	AATACTCTCGAGACGGACAATAAAACG	*K.lactis HHO1* plasmid
Hho1_SacIIR	ATTCGACCGCGGATTAACTTTATTTCTTGGAC	*K.lactis HHO1* plasmid
Isw1_XhoIF	ACTTTGCTCGAGTCCAAATTAATACACAATATAG	*K.lactis ISW1* plasmid
Isw1_SacIIR	AAAACACCGCGGCCACTCTACCATTCTTGC	*K.lactis ISW1* plasmid
Ioc3_XhoIF	TCAAGTCTCGAGGACTGCCAGCTTCTCC	*K.lactis IOC3* plasmid
Ioc3_SacIIR	GCTTTTCCGCGGAAAGAGTGAACGGCCATCC	*K.lactis IOC3* plasmid
SChd1::KChd1 5’	ATTCAAAGCAGAACCTTTTCTAATTTAATTCTCAC TTATAATG*GTACAGGATTTACCAG*	*CHD1* swap
SChd1::KChd1 3’	GAAGGAACAATGGAAAATGTGGTGAAGAAAAATT GTTTCA*ATGTTCAGAGGCTTTGATT*	*CHD1* swap

Tabel V. 1: Primers used for Chd1 deletion and “swap” strain construction.

HHO1, *ISW1*, and *CHD1* were deleted from *S.cerevisiae* by transformation with PCR products generated using pRS416 as a template for *URA3*. *IOC3* was deleted with *KANR* PCR from pFA6:HA:KanMX6 with the above primers. Sequences for amplifying *URA3* and *KANR* are italicized. *HHO1*, *ISW1*, and *IOC3* were amplified with the above primers from *K.lactis* genomic DNA and cut with the appropriate restriction enzyme (restriction enzymes sites are underlined) and ligated into cut vector (pRS415 for *ISW1* and *HHO1* and pRS413 for *IOC3*). *K.lactis CHD1* coding sequence was amplified from *K.lactis* genomic DNA with *S.cerevisiae CHD1*-tailed primers; *K.lactis CHD1* sequence is italicized.

Primer	Sequence	Purpose
SChd1_BamHI F	ATCCCAGGATCCTTATCACATCCCA	Clone *CHD1*
SChd1_SacII R	ATTACTCCGCGGTGCCCTATAGTACACAC	Clone *CHD1*
S::KChd1_swapN 5'	AAAGCAGAACCTTTTCTAATTTAATTCTCA CTTATAATGG*TACAGGATTTACCAGACG*	Chimera: *K.lac* (1-351)
S::KChd1_swapN 3'	AATTACTGGAATATTGTGGGAGGATCTTA GAGTTTTCTCT*GTTTTGGAACTCTTTTACC*	Chimera: *K.lac* (1-351)
S::KChd1_swapC 5'	TAAAGCTGATATAGATTGGGATGATATCAT TCCAGAAGAAG*AACTCAAGAAACTAAAGG*	Chimera: *K.lac* (954-1525)
S::KChd1_swapC 3'	GGGGAAGGAACAATGGAAAATGTGGTGAA GAAAAATTGTTT*TAATGTTCAGAGGCTTTG*	Chimera: *K.lac* (954-1525)
S::KChd1_swapA 5'	GAAATTGGCACCTGAACAAGTGAAACATT TTCAAAACAGA*ACCAATTCCAAGATAATGC C*	Chimera: *K.lac* (352-953)
S::KChd1_swapA 3'	CCTTGCGTTTCTGCTCTTCATCTTGGAGTT TTTTTAGTTC*ATCTTCGGGGATAATGTCA*	Chimera: *K.lac* (352-953)
S::KChd1_chromoswap 5'	TGTGCATGAAGCATCTGCCAATCCTCAAC CAGAGGACTTC*CATAGCATCGACATCGTT G*	Chimera: *K.lac* (177-351)
S::KChd1_chromoswap 5'b	TGTTATCAATCACAGACTAAAGACATCTTT GGAAGAAGGA*GGATCGACGAAGAATCGT AG*	Chimera: *K.lac* (195-351)
SChd1 100up 5'	ACTAACGACAAAGTTTCTCAAAGG	*S.cerevisiae CHD1*
K::SChd1_unsNswap 3’	*CTTCCTTCAAACGATGATCGACAACGATG TCGATGCTATG*GAAGTCCTCTGGTTGAGG	Chimera: *K.lac* (1-176)
K::SChd1_unsNswap 3’b	*TGATGGAATCCACATCCCATCTACGATTCT TCGTCGATCC*TCCTTCTTCCAAAGATGTC	Chimera: *K.lac* (1-194)

Tabel V. 2: Primers used to generate chimeric *CHD1*.

Restriction enzyme (sequence underlined) tailed primers were used to amplify from *S.cerevisiae* (BY4741) genomic DNA and clone *CHD1* into cut pRS415 vector. Yeast homologous recombination was used on this plasmid to replace *S.cerevisiae CHD1* sequence with *K.lactis*’ by digesting the Chd1 plasmid and transforming along with the *K.lactis* generated PCR products for swapN, swapC, and swapA. *K.lactis* chromodomains were swapped by homologous recombination of PCR product amplified from *K.lactis* genomic DNA using S::K_chromoswap 5’ or 5’b and S::KChd1_swapN 3’ in the *S.cerevisiae CHD1* plasmid. PCR product amplified with SChd1 100up 5’ and K::SChd1_unsNswap 3’ or 3’b was recombined into the N-terminal chimeric *CHD1* plasmid to generate plasmids containing the first 176 or 194 amino acids of *K.lactis CHD1* in place of the *S.cerevisiae* sequence. These plasmids were linearized and transformed into the deletion strain to integrate the chimerae.

Yeast Growth:

Hho1 and Isw1Ioc3 strains were grown in Synthetic Dextrose +URA3 media to maintain plasmids. Chd1 integrated strains were grown in YPD media. All yeast were grown in 200mL of media at 30°C with 220rpm shaking in an Innova 44 incubator overnight until about 0.5 OD_{600}.

Mononucleosomal DNA Isolation:

Yeast cultures (200mL) were fixed in 1.85% formaldehyde at 30°C, shaking, for 30 minutes. Yeast cultures were pelleted, washed, and resusupended in 1.2mL of cell breaking buffer (10mM Tris, pH 7.4, 20% glycerol) with Sigma protease inhibitors and split into 2 screw cap tubes with approximately 0.6mL of 0.5mm zirconia/silica beads (Biospec Products). The cell wall was broken in a Mini-Beadbeater-96 (Biospec Products) in a cold magnetic rack with two 2 minutes on the bead beater and an additional 1 minute and 17 seconds on the bead beater, each of these was separated by 1 minute and 15 seconds on ice. The cells were checked under a microscope to ensure that most were broken; the lysed cells were separated from the beads by spinning them through a puncture in the bottom of the tubes into 5mL tubes. The cells were spun at maximum speed in a 4°C eppendorf centrifuge and the supernatant was discarded. The cells were resuspended up to about 2.4mL with NP buffer (50mM sodium chloride, 10mM Tris pH7.4, 5mM magnesium chloride, 1mM calcium chloride, 0.5mM spermidine,

1μL/mL β-mercaptoethanol, and 0.01% NP-40) and 0.6mL of cells was aliquoted to four tubes with micrococcal nuclease aliquoted in the lid. The cells were mixed with the enzyme and incubated at 37°C for 20 minutes, after which 150μL of STOP buffer (50% sodium dodecylsulfate, 0.05M ethylenediaminetetraacetic acid) and 5μL 20mg/mL proteinase K and the digestions were incubated at 65°C overnight to remove the histone proteins. Mononucleosomal DNA was purified by a PCI extraction and precipitation with 0.3M sodium acetate in 2-propanol and then dissolved in 60μL of 1X NEB buffer 2. The DNA was digested with 2μL of 20mg/mL RNase solution (Sigma) at 37°C for 1 hour. Digestion ladders were assessed by running 5μL on a 2% agarose gel. Digestions were chosen to have mostly mononucleosomal-sized DNA with a little dinucleosomal and a hint of trinucleosomal DNA. Half (25μL) of the appropriate digestion was treated with calf intestinal phosphatase (0.75μL) at 37°C for 45 minutes. The mononucleosomal band was gel purified from a 1.8% agarose gel, using Freeze N' Squeeze columns (BioRad), and was PCI extracted and ethanol precipitated overnight.

Deep Sequencing Library Preparation:

About 250ng of mononucleosomal DNA was end-cleaned with END-it (Epicentre) in a 40μL reaction. Agencourt AMPure XP beads (Beckman Coulter) beads were used to clean-up the reaction with 1.8X beads and washed once with 70% ethanol. The beads were resuspended with 21.25μL water and the DNA was

then A-tailed with Klenow exo- DNA Polymerase (Epicentre) in a 25μL reaction on the beads. This reaction was cleaned up on the beads by adding 1.8X ABR buffer (15% PEG, 2.5M sodium chloride) and washed with 70% ethanol. The beads were resuspended with 11.75μL water and 0.6μL 10nM barcoded adapters was added; a 15μL ligation reaction was performed at room temperature for 1 hour and the reaction mixture was made up to 25μL and the ligation was continued at 16°C overnight. The ligation reaction was cleaned up on the beads by adding 1.3X ABR buffer, washed twice with 70% ethanol, and eluted with 39μL water. Two-thirds of the ligated DNA was used for two 25μL Pfx PCR reaction, using 10 and 12 cycles of the PCR extensions. A portion of the PCR reactions (4μL) was run on a gel to check the presence and size of the product. A final PCR with the remaining one-third of the ligated material was repeated with the best cycling conditions for each sample. PCR reactions were pooled and mixed with 1.3X AMPure beads, washed on the beads two times with 70% ethanol, and eluted in 20μL water. One tenth of the purified library was used to quality check the library by StrataCloning, which requires an initial A-tailing reaction with Taq DNA Polymerase. KAPA Library Quantification Kit was used on a serial dilution of 1μL of each library to estimate the relative concentrations, and the libraries were mixed 1:1 and sent to the UMass deep sequencing core for either Hi-Seq or GAII (N-terminal Chd1 chimerae) sequencing.

Data Analysis:

Raw, fastq files were separated by barcode, using Novobarcode. Sequences were aligned to the SacCer3 genome using bowtie2 with its defaults, and filtered to keep only the uniquely mapped reads, separating reads into Watson and Crick aligned. In order to make TSS alignments, aligned reads were counted using bedtools genomecoverage for every basepair. The cross-correlation between Watson and Crick reads was used to infer fragment length and reads were extended to the fragment length and recounted for the full fragment, using bedtools genomecoverage. The genome average base pair count was averaged to one and alignments were made using 500 base pairs upstream of the edge of the +1 nucleosome (as defined in Tsankov et al., 2010) (Tsankov et al., 2010) to the end of the gene.

In order to call nucleosome positions, the first base pair of each reads was added up and the Watson and Crick reads were combined in order, first by chromosome then base pair, generating a tab delimited file with chromosome, base pair, F/R, and count. This file was then used in a Template Filtering algorithm with 7 nucleosome templates (developed in Weiner et al., 2010) to call nucleosome positions genome wide. In order to assign nucleosomes to genes and designate the +1 and -1 nucleosomes, a homemade perl script scanned nucleosome calls up to 500 base pairs upstream of the coding region to find the nucleosome depleted region (a linker length greater than 100 base pairs) upstream of the

gene, the +1 and -1 nucleosomes were designated as the flanking nucleosomes, and the positions of the genic nucleosomes were assigned to each gene. The average distance between the genic nucleosomes was determined for each gene, as was the distance between the centre of the +1 and +3 and +3 and +5 nucleosomes.

The RNA Polymerase II occupancies defined in the unstressed wild type strain in (Kim et al., 2010) were used to separate genes by Pol2 occupancy. The wild type and chd1 mutant strain turnovers were used from (Radman-Livaja et al., 2012) to separate genes by average coding region turnover as well as the difference in 3' turnover. Data from N-ChIP of Chd1 performed in (Zentner et al., 2013) was used to separate genes by Chd1 immunoprecipitation levels.

CHAPTER VI: Discussion

Nucleosomes restrict access to the underlying DNA sequence, and hence contribute to regulating processes that occur on DNA, including gene expression. In order for transcription to initiate, transcription factors compete with nucleosomes for binding to DNA. Furthermore, RNA Polymerase contends with the nucleosomal barrier to transcription initiation and elongation. The packaging of the eukaryotic genome must allow for regulated access of transcription factors and the transcriptional apparatus to DNA. The blocking of transcription factor access to a chromatin template is partially relieved by the surprisingly characteristic nucleosome depletion at promoters that was discovered in genome-wide microccocal nuclease mapping of chromatin (Yuan et al., 2005). Indeed, nucleosome organization at promoters tends to be the most open at more highly expressed, growth or housekeeping genes as compared to more lowly expressed, stress responsive or tissue-specific genes (Ioshikhes et al., 2006; Schones et al., 2008; Tirosh and Barkai, 2008; Yuan et al., 2005).

Given the strong conservation of genic chromatin architecture and its suspected contributions to gene expression, discovering the factors involved in establishing nucleosome positions surrounding genes has been a fundamental pursuit in the field. Sequence preference plays a role in nucleosome depletion (Kaplan et al., 2009), but the extent to which sequence determines chromatin architecture is debated (Zhang et al., 2009). NDR generation via sequence-

based exclusion of nucleosomes is compatible with constitutive expression; however, the utility of sequence-based nucleosome depletion is less clear for regulated NDR formation and gene expression. While yeast promoters often contain sequences that prevent nucleosome formation, human promoters lack these “programmed” NDRs; this at least partially reflects the scarcity of unexpressed genes in yeast relative to humans. Clearly, inducible genes would benefit from *trans* regulated promoter access, as is the case for promoters such as *PHO5* (Almer et al., 1986).

Overall, this body of work examines the relative contributions of *cis* and *trans* factors to nucleosome positioning in an *in vivo,* hybrid system. Hybrid approaches are useful in distinguishing between *cis* and *trans* effects, and previous studies in either haploid segregants of *S.cerevisiae* strains or diploids generated from closely related yeast species have linked gene expression and promoter chromatin to *trans* and *cis* determinants (Brem et al., 2002; Tirosh et al., 2010). Using YACs allows us to make comparisons between more divergent species than was possible with prior hybrid approaches which depended on successful mating, thus providing us with greater power to observe *trans* regulation. Using our YAC system we conclusively show that while sequence preference (or conserved trans-acting proteins) plays a role in promoter chromatin structure – based on the maintenance of NDRs on the same sequence in different environments – the preponderance of nucleosomes are positioned by *trans*-acting factors. This is particularly evident in the generation of larger genic

internucleosome spacing by a *K.lactis* factor, relative to the spacing generated in *S.cerevisiae*, on the same DNA sequence. In fact, even the +1 nucleosome and the NDR, which present the boundary against which downstream nucleosomes are packaged, are variable in different nuclear environments.

While we do confirm that sequence biases play a role in *S.cerevisiae* NDRs, this is not a consistent theme across other organisms, and in fact, sequence-based nucleosome depletion is not even strongly maintained across Hemiascomycete yeast, as *D.hansenii* sequence is less apt to inform nucleosome depletion. This is perhaps linked to differential requirements to regulate access to the promoter region, as most *S.cerevisiae* genes are expressed and do not require tissue-specificity or developmental regulation as in metazoans. Nevertheless, we clearly see a role for *trans* factors in clearing yeast promoters. A role for *trans* acting factors, such as GRFs and remodelers, in establishing NDRs would allow for another tier of control in gene expression, by regulating the access of the preinitiation complex to DNA.

In Chapter III, we confirm that *trans* regulators are involved in and sometimes required for the generation of many NDRs. While this is less apparent in species like *S.cerevisiae* and *K.lactis*, where promoter sequence content is fairly predictive of nucleosome depletion, others, such as *D.hansenii*, appear to require the input of *trans* factors for at least half of their NDRs. In fact, this approximation, based on the number of YAC borne promoters that gain

nucleosome occupancy, is likely an underestimate due to the presence of conserved *trans* acting factors between yeast species. Human genes which encode nucleosome occupancy over their promoters via high GC content (Tillo et al., 2010) are even more likely to be dependent on *trans* acting factors for an open and active state.

Trans acting factors could participate in nucleosome depletion through remodeling chromatin or competing with histones for a binding site. Given the association of YAC-specific neoNDRs with YAC-specific transcripts, it is likely that transcription factors are frequently involved in NDR generation through the latter mechanism. Indeed, we show in Chapter IV that transcription factor footprints are enriched in both native and YAC-specific NDRs, and are absent in conditions where the sequence is occupied by nucleosomes. The simplest explanation is that these transcription factors compete with histones; however, we cannot rule out the contribution of chromatin remodelers at these locations, and in fact we believe that recruitment of chromatin remodelers may be an important property of some sequence-specific binding proteins. This could be what distinguishes Cbf1 as a General Regulatory Factor in *D.hansenii*, as we show in Chapter IV that the *D.hansenii* protein itself does not inherently contain nucleosome organization activities. Interestingly, in the absence of Cbf1, nucleosome occupancy is increased at loci where it binds in *S.cerevisiae*, proving that transcription factor binding can compete with histones, or recruit histone-evicting activities, to generate NDRs. Studies to find what feature(s) of

D.hansenii Cbf1 differ from that of *S.cerevisiae* could reveal what defines GRF identity and function.

Binding of transcription factors is not the only event in gene expression that is capable of affecting nucleosome positions. In Chapter II, we address the ability of RNA polymerase II to disrupt nucleosomes. We see that even the +1 nucleosome, which has been argued to be the most well-positioned by sequence (Brogaard et al., 2012), is shifted downstream upon transcriptional inactivation. This retrograde movement of genic nucleosomes during transcription is consistent with RNA polymerase passing nucleosomes backwards as it transits the gene, as has been seen *in vitro* (Studitsky et al., 1994, 1997). Furthermore, the phasing of nucleosomal arrays decays without active transcription. Additionally, in Chapter III we saw the de novo generation of well-positioned arrays with the advent of transcription from the heterologous sequence of *D.hansenii* YACs. This substantiates the involvement of transcription elongation and associated factors in helping to form well-positioned arrays.

The genome-wide positioning of the TSS just within the borders of the yeast +1 nucleosome that we show in Chapter II and others have also documented (Yuan et al., 2005) was initially puzzling; how does transcription initiation occur from a location that is partially obscured by a well-positioned nucleosome? This question has lead to a related one: what establishes the relationship between the positioning of the TSS and +1 nucleosome? This could

result from either the +1 nucleosome restricting TSS selection or PIC assembly stabilizing formation of the adjacent +1 nucleosome. In Chapter III we continue to observe an intimate relationship between transcription initiation and nucleosome organization, where a species-specific relationship is maintained despite shifts in TSS selection or *de novo* generation of functional promoters. We determine that a *trans* acting factor is establishing this relationship for each species, and we see that the TSS and +1 nucleosome positions are set by a *trans* acting factor or factors and not by sequence. While transcription clearly plays a role in shaping chromatin structure both at the +1 nucleosome and downstream, the relationship between chromatin and transcription, particularly the stereotyped TSS to +1 link (discussed in Appendix AI,) does not appear to be a simple one. The simplest model to explain the link between transcription and chromatin organization is that sequence and transcription factor binding coordinate to generate NDRs, which stabilize the adjacent +1 nucleosome and allow for PIC assembly; transcription then serves to fine tune the position of the +1 nucleosome and to generate phased nucleosomal arrays. Evidently sequence and transcription as well as other factors cooperate to establish *in vivo* nucleosome positions, where sequence primarily plays a role in promoter nucleosome depletion, and transcription and other *trans*-acting factors are involved in the formation of chromatin structure at the promoter as well as in the gene body.

We saw that the distance between nucleosomes is measured by a *trans* acting factor that appears to be associated with RNA polymerase II. The fifth and

final chapter expands on the finding from Chapter III that nucleosome spacing is guided by a *trans* acting factor or “molecular ruler”. This means that an evolutionary difference in a *trans* factor, which actively spaces nucleosomal arrays, causes an increased measurement between nucleosomes; this has the benefit that deleting candidate factors from *S.cerevisiae* and complementing the deletion with the *K.lactis* copy would lead to increased spacing in *S.cerevisiae*. In Chapter V, we generate complementation strains of candidates for nucleosome spacing and map nucleosomes. While neither the *K.lactis* copies of the linker histone, Hho1, nor the Isw1a complex caused an increase in nucleosome spacing in *S.cerevisiae*, replacing *S.cerevisiae* Chd1 with the *K.lactis* copy did result in increased nucleosome repeat length. Consistent with reports of Chd1 interacting with elongation factors (Biswas et al., 2007; Simic et al., 2003), genes with the highest RNA polymerase occupancy exhibit a greater increase in internucleosome distance in the presence of *K.lactis* Chd1 than do genes with lower polymerase occupancy. However, high levels of nucleosome turnover do not appear to be necessary for Chd1 spacing activity. This could mean that Chd1’s activity in the gene body is dependent on its association with elongation factors and does not require transcription-associated turnover as has been postulated (Zentner et al., 2013), or that Chd1 spacing activity does require greater exposure of DNA than the linker, but that this comes from partial unwrapping of the nucleosome rather than wholesale turnover.

While the increased internucleosomal length generated by *K.lactis* Chd1 is distributed throughout the protein, the first 176 amino acids, upstream of the chromodomains, are the smallest domain that can generate increased linker length when swapped between species. Little is known about the extreme N-terminus of the protein as the first 117 amino acids are typically omitted in biochemical assays. This is particularly interesting for a few reasons. Firstly, these residues that are uncharacterized *in vitro* appear to have *in vivo* significance. Secondly, *in vitro* assays have reported that recombinant Chd1 missing the N-terminus is capable of *in vitro* centering activity (Hauk et al., 2010), indicating that while these residues lead to some difference in the measurement between the species they are not required for sliding and spacing of nucleosomes. Potentially, the N-terminus may modulate the folding of the chromodomains against the ATPase lobes of Chd1 and sensing of the histone H4 tail, or it may influence the stability and thus protein expression levels of these species' Chd1. In this case the nucleosome linker length differences between species would not result from a change in the activity or physical measurement of the protein, but from the amount of active protein available across the gene body. Protein abundance, stability, and localization studies are ongoing, and should shed light on the possibility of the N-terminus of *K.lactis* Chd1 altering spacing by this latter mechanism.

The generation of an *S.cerevisiae* strain that differs only in the copy of Chd1, leading to an increase in internucleosome spacing, presents an interesting

opportunity to examine the importance of nucleosomal array spacing in the regulation of the genome. Does the space between nucleosomes have a functional consequence on gene expression? Does it alter how the elongating polymerase deals with the nucleosomal barrier? Does it change longer range chromatin interactions and packaging? The functional consequences of nucleosomal spacing can now be determined with this new tool.

Chromatin has a great capacity to regulate gene activation by integrating multiple layers of information. It provides robustness to maintain gene expression levels by allowing access to open, highly expressed genes and hindering assembly of the transcriptional machinery at closed, unexpressed genes. Additionally, the responsiveness of chromatin to RNA polymerase transit may help to maintain a genic chromatin signature consistent with gene expression, thus perpetuating a chicken-and-egg cycle. The multi-faceted regulation of chromatin structure allows for the further tuning of gene expression by more than just transcription factor presence. By integrating sequence biases, transcription factor binding sites, and recruitment of chromatin remodelers, the cell can enact myriad responses. This can be appreciated in the tuning of expression levels from promoters constructed with different TF binding sites and anti-nucleosomal sequences (Raveh-Sadka et al., 2012). Preventing nucleosomes from obstructing TF binding sites by the incorporation of anti-nucleosomal sequences should aid in the robust expression of genes. On the other hand, the requirement of *trans* acting factors to antagonize nucleosome occupancy can lead to variable

responses depending on the expression level and recruitment of the *trans* factor(s). Our results conclusively establish a large contribution by *trans* regulators to genic chromatin architecture and thus indicate that chromatin can evolve and be modulated with alteration of even an individual regulator.

CHAPTER AI: Defining the order of the relationship between the locations of the TSS and +1 nucleosome.

Introduction

The surprisingly well defined location of the *S.cerevisiae* transcription start site (TSS) about 10bp into the +1 nucleosome was an early revelation of genome wide nucleosome mapping (Yuan et al., 2005). In contrast to larger eukaryotes, in *S.cerevisiae* TSSs are not located at a fixed distance from the TATA box (Chen and Struhl, 1985; Zhang and Dietrich, 2005). This has lead to the so-called “scanning” model for transcription initiation, in which RNA Polymerase scans DNA starting from about 20bp downstream of the TATA box (Giardina and Lis, 1993) until it reaches an appropriate initiator sequence (Inr); in this model, the +1 nucleosome could serve to restrict the movement of polymerase and thereby help determine the position of the TSS. In fact, ChIP exo data for the preinitiation complex in yeast showed that the TBP associated factor, Taf1, was well-positioned relative to the +1 nucleosome rather than a poorly conserved TATA box, suggesting a role for the +1 nucleosome in restraining the TSS selection at these TATA-less promoter (Rhee and Pugh, 2012). On the other hand, transcription can also play a role in shaping the chromatin architecture; *in vitro* transcription assays have shown the transit of polymerases to move nucleosome upstream or result in dimer or whole octamer eviction (Bintu et al., 2011; Kulaeva et al., 2010). Additionally, in Chapter II we have observed the

shifting of +1 nucleosomes into the gene body, upon loss of RNA Polymerase activity (Weiner et al., 2010). TFIIB appears to play a role in defining the transcription start site, as *S.pombe* TFIIB and RNA Polymerase II generate start site selection typical of *S.pombe* in an *S.cerevisiae* cell extract system (Li et al., 1994). Additionally, mutations in the TFIIB "B reader" domain, such as E62K, shift start site selection to downstream Inrs, presumably by modulating the threading of the template strand through the template tunnel in RNA Polymerase II-TFIIB structure (Kostrewa et al., 2009). These contrasting observations raise the question of whether the stereotyped relationship between +1 nucleosomes and TSSs arises because the preinitiation complex helps stabilize nucleosome formation nearby, or because positioned nucleosomes help constrain Pol2 scanning and thereby influence TSS selection. As an approach to disentangle the order of TSS selection and +1 nucleosome establishment, we mapped TSSs and nucleosomes in a yeast strain carrying a mutation in TFIIB known to affect transcription start site selection at some promoters, *sua7E62K* (Pinto et al., 1994).

Results and Discussion

TSS mapping in the *sua7* mutant revealed that this mutation does not globally induce downstream shifts in start site selection at all genes; in fact, TSS changes in this mutant are rarely a simple shift in location (Fig. AI-1A). It is possible that detectable TSS shifts are confounded by decreased stability in

some transcripts arising from aberrant TSSs. Additionally TSS selection is influenced by local sequence as well as the copy of TFIIB. If the location of the TSS is set and then transcription from this spot dictates the position of the +1 nucleosome, we would expect a shift in the +1 nucleosome at genes with altered TSS selection. In general we observed that most genes do not experience a shifted +1 nucleosome in the *sua7E62K* mutant (Fig. AI-1B) even in many cases where TSS selection is shifted (Fig. AI-2B). This may represent an instance where the TFIIB mutant shifts the minimal distance from the TSS to a downstream sequence that is not usually restricted by the wild type nucleosome architecture.

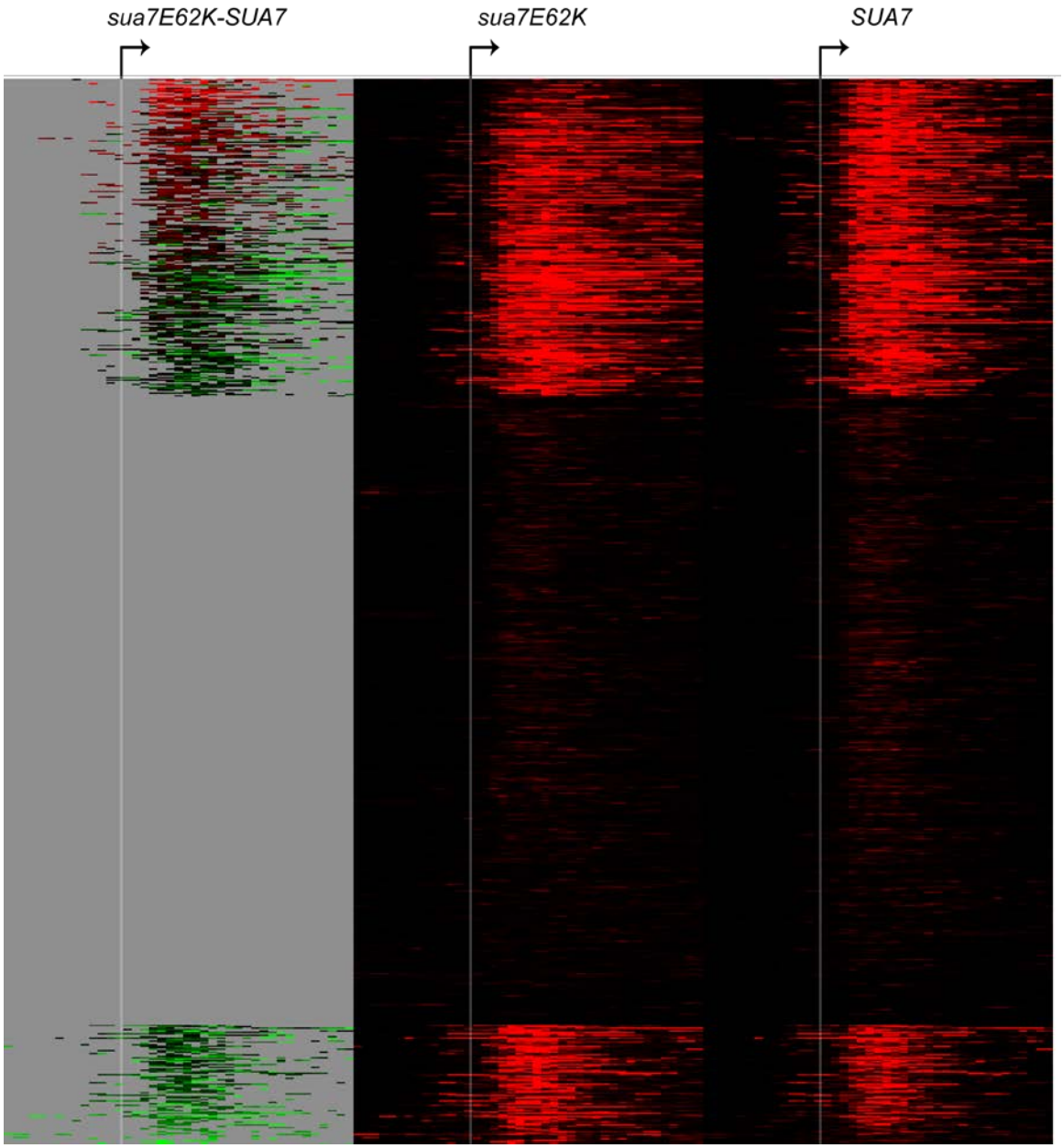


Figure AI. 1: The *sua7E62K* mutant does not shift TSS selection downstream genome-wide.

Mapping of TSSs in *sua7E62K* (centre panel) and wild type (right panel) yeast shows that TSS selection is not simply shifted downstream in the mutant as has been seen at individual loci, and in some cases leads to tightening up of TSSs.

We do observe some cases of concurrent shifts in TSS utilization and +1 nucleosome location, but also genes where alterations in transcription and nucleosome architecture are discordant (Fig. AI-2A-E). At the extremity, Fig. AI-2C shows the concurrent gain of an NDR and increase in TSS abundance at this gene in the *sua7E62K* mutant. Alterations in chromatin structure in the *sua7E62K* mutant, which is not expected to have any direct effect on nucleosomes, support a role for transcription or PIC formation in nucleosome architecture. However, we cannot establish any particular cause and effect relationship between TSS and +1 nucleosome position from this data. This negative result is inconclusive, as these data represent population averages – subtle changes in TSS locations are easy to find due to the digital nature of these data, while nucleosome peak locations are "blurred" by MNase chewing. This is potentially illustrated for YOR198C, shown in Fig. AI-2D, where the low occupancy +1 nucleosome loses even more nucleosome occupancy. This appears to be accompanied by a slight shift in preference for a downstream TSS peak that is present in both strains. The downstream TSS could thus represent a preferred sequence that is masked in a greater population of the wild type culture than in the *sua7E62K* mutant population. However there are many instances where the differential location of the +1 nucleosome is not accompanied by any alteration in TSS selection (Fig. AI- 2E). It is possible that shifts in the +1 nucleosome are concordant with shifts in TSS, but that transcripts originating from certain loci are less stable and cannot

be detected by TSS mapping of steady state transcripts. Furthermore, shifts in nucleosomes could conceivably result from indirect effects via alteration of expression of any number of chromatin regulators.

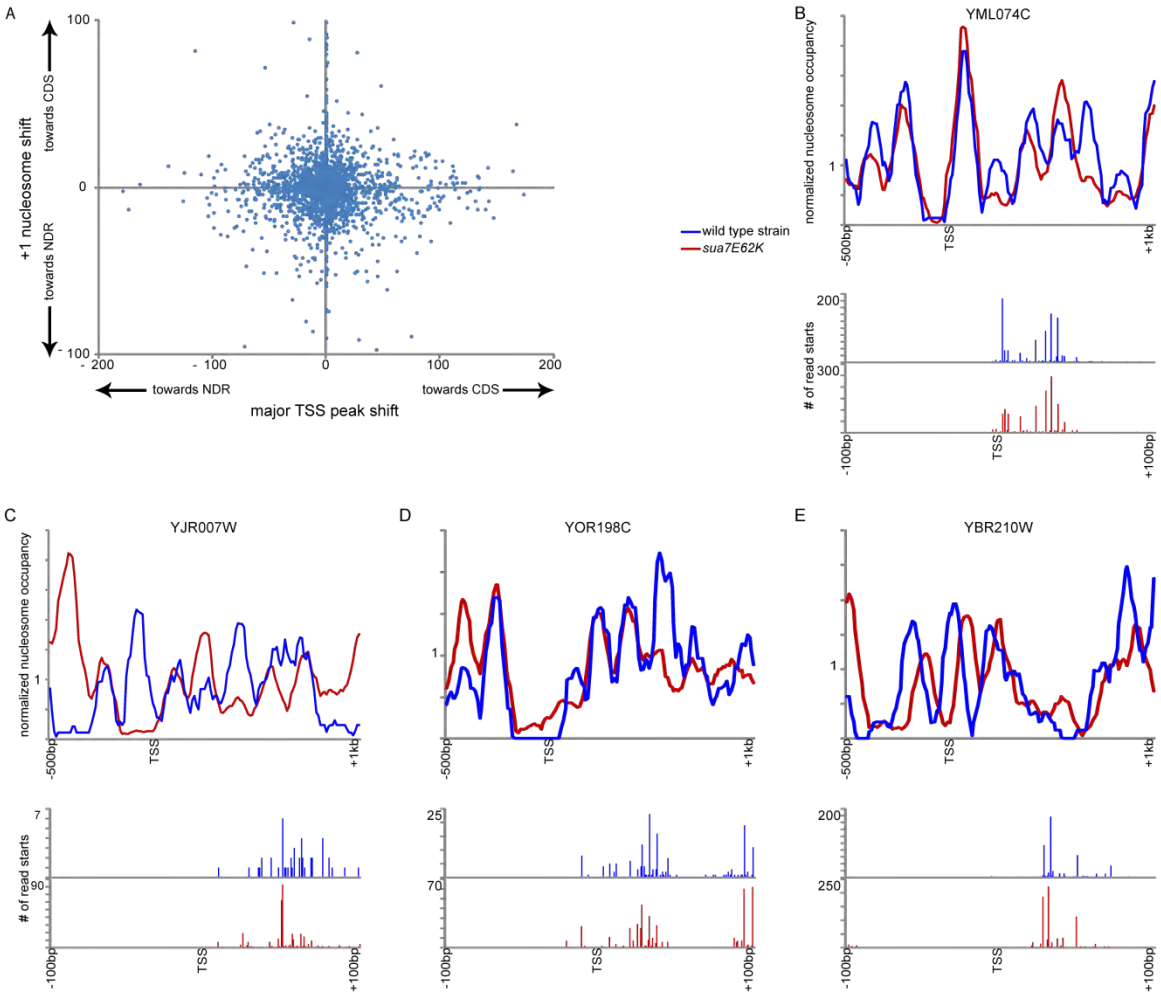


Figure AI. 2: Nucleosome positioning is affected in the *sua7E62K* mutant.

A) A scatterplot of shifts in the major TSS peak and shifts in the +1 nucleosome shows that the +1 nucleosome and TSS selection are not correlated.
B) YML074C represents a gene where TSS selection shits downstream but is not accompanied by alterations in +1 nucleosome positioning.
C) YJR007W shows a striking case where the loss of a positioned nucleosome to generate an NDR in the TFIIB mutant is associated with a large increase in transcription.
D) YOR198C shows a decrease in partial nucleosome occupancy along with an increase in downstream TSS selection.
E) YBR210W is a case where the TFIIB mutation does not manifest in TSS selection but does result in nucleosome shifts.

Materials and Methods

Yeast strains and growth:

The YSB143 *SUA7* shuffle (genotype: MATa *ura3-52 leu2-3,112 his3d200 sua7ΔLEU2* [pRS313-*SUA7* CEN/ARS *HIS3* AmpR]) and YSB176 *sua7E62K* (genotype: MATa *ura3-52 leu2-3,112 his3d200 sua7ΔLEU2* [pRS313-*sua7-35 (E62K)* CEN/ARS *HIS3*]) strains were kindly provided by the Buratowski lab. Yeast (2 cultures each) were grown in 450mL YPD (1% yeast extract, 2% bacto-peptone, 2% dextrose) cultures overnight at 28°C overnight shaking at 220rpm in an Innova 44 incubator to an OD_{600} of ~0.5. One set of cultures was pelleted, snap frozen, and stored at -80°C for later RNA extraction and TSS mapping, while the other was fixed with 1.85% formaldehyde of nucleosome mapping via micrococcal nuclease sequencing.

RNA Isolation:

RNA was extracted by vortexing cell pellets in 6mL Trizol and ~2.5µg beads for 5 minutes, followed by chloroform extraction and 2-propanol precipitation of the aqueous phase. About 400µg of resuspended RNA was cleaned and DNaseI treated on Qiagen RNAeasy columns. polyA RNA was enriched on NEB oligo dT magnetic beads with two rounds of selection by adding 1/3 the volume of 2M Lithium Chloride to the first eluant and reapplying to oligo dT beads. Lithium was removed from the RNA via Qiagen RNeasy MinElute and ~2µg was used for TSS library construction.

TSS Library Construction (performed by Ting Ni in the Zhu Lab as in (Ni et al., 2010)):

RNA was treated with 2.4U of Bacterial Alkaline Phosphatase (0.4U/μL, Takara) for 40 minutes at 37°C in a 100μL reaction with RNase inhibitors. The reaction was cleaned up by a phenol:chloroform and then chloroform extraction and precipitated with GlycoBlue. BAP-treated RNA was treated with 20U Tobacco Acid Pyrophosphatase (Epicentre, 10U/μL) in a 100μL reaction with RNase inhibitors to remove the mRNA cap and leave a ligatable 5' phosphate. This reaction was similarly cleaned up. RNA was then ligated to a chimeric linker in a 100μL T4 RNA ligation reaction with 25% PEG 8000, 200U of T4 RNA Ligase (NEB), and RNase inhibitor. The ligation was cleaned up via phenol:chloroform and chloroform extraction and excess linker was removed with RNeasey MinElute kit. Reverse transcription was performed in a 40μL reaction, first incubating 20μL of RNA with 20pmoles of dNTPs and 20pmoles of Mme_RTN6 primer at 65°C for 5 minutes and snap chilling; the final reaction contained 5mM DTT, 6ng/μL actinomycin D (freshly diluted) and 2μL SSIII reverse transcriptase in 1X FS buffer (Invitrogen). RNaseH (2U, Invitrogen) was used to remove the RNA strand and the cDNA was purified in a Zymo Research clean n' concentration kit, eluting in 20μL water. A quality control end-point PCR was performed with 0.25μL of cDNA with primers for the sequences added in the ligation and RT steps to check for the presence of product. A low cycle 50μL PCR in 1X HF buffer (Finnzymes) containing 500nM of each primer for the added

sequence and 0.2mM dNTPs was performed with 0.5µL Phusion (Finnzymes). 5µL ExoI exonuclease was added to the PCR, incubating at 37°C for 45 minutes and inactivating at 80°C for 20 minutes to remove excess primers. The amplified DNA was purified with Zymo Research DNA clean n' concentrate and eluted in 11µL water. The product was circularized, using 0.9µL 10µM TSS_Col3_short oligo and 1.5µL 5U/µL Ampligase (Epicentre), after combining 10µL of the RT-PCR with 3µL 10X Ampligase buffer, 1.5µL Optikinase (5U/µL, USB), 1.5µL 100mM ATP, 0.3µM 100mM DTT, and 11.3 µL water and incubating at 37°C for 30 minutes and heat inactivating the Optikinase at 95°C for 2 minutes. After adding Ampligase and the oligo, the circularization proceeded with 5 cycles (95°C, 30sec; 68°C, 2min; 55°C, 1min; 60°C, 5min) and then 5 cycles (95°C, 30sec; 65°C, 2min; 55°C, 1min; 60°C, 5min). Oligo and linear products were removed with the addition of 3µL ExoI (NEB) and 0.6µL ExoIII (NEB) and incubation at 37°C for 45min and 80°C for 20min. Four reaction of rolling circle amplification were performed with 2µL circularized product in a 20µL reaction containing 1mM dNTPs, 0,2mg/mL BSA (NEB), 1X Phi29 buffer (NEB), 10µM N6T2 oligo, 10U Phi29 DNA polymerase, UV-treated (NEB), and 2µL DMSO. RCA was performed at 10°C for 10min, 28°C, for 16hours, and 65°C, for 10min. The circular RCA product was confirmed with an XhoI digestion (using 2µL of the total combined reactions). RCA products were purified with phenol:chloroform and then chloroform extraction followed by ethanol precipitation. Approximately 16µg of RCA product was digested with 20U of MmeI in a 200µL reaction

containing 1X NEB buffer 4 and 50μM S-Adenosylmethionine at 37°C for 30min. The digestion was loaded on a 6% acrylamide gel and separated at 100V for 2hours and then 200V for 30min. The 94bp band was cut from the gel and eluted in 400μL gel elution buffer (0.1%SDS, 0.32M sodium chloride, 10mM magnesium acetate) rotating at 4°C overnight. The eluant was extracted with phenol:chloroform, then chloroform, and was ethanol precipitated with GlycoBlue. The digested product, containing the TSS and a downstream tag flanking a central sequence from the ligation and PCR oligos, was ligated in a 5μL T4 DNA Ligation reaction (NEB) with 16% PEG 8000 and 4000U T4 DNA Ligase (NEB) to 375nM Illumina paired-end primers (PE2_AN2 and PE2_BN2) at 16°C overnight. The ligation reaction was phenol:chloroform and chloroform extracted and ethanol precipitated with GlycoBlue and subsequently separated on an 8% acrylamide gel at 1W for 1 hour. The 180bp band was cut out and eluted in 400μL gel elution buffer at 4°C overnight. This was purified in a Zymo Research DNA clean n' concentrate kit and a final low cycle PCR was performed in a 50μL 1X HF buffer (Finnzymes) reaction with 50nM each PE2_AFN2_short and PE2_BRN2_short, 500nM each PE2_A_short and PE2_B_short oligos, 0.2mM dNTPs, and 0.5μL Phusion DNA Polymerase (Finnzymes). PCR cycles were as follows: 98°C, 30sec, 2cycles (98°C, 10sec; 66°C, 10sec; 72°C, 30sec), 10cycles (98°C, 10sec; 69°C, 10sec; 72°C, 30sec), and 72°C for 10min. The PCR reaction was purified with Zymo Research DNA clean n' concentrate kit, and this was used for paired-end Illumina deep sequencing.

Oligo	Sequence	Used
Chimeric linker	CTCAAGCTTCTAACGATGTACGCTCG rArGrUrCrCrArArC	Ligation to TSS
Mme_RTN6	GCGGCTGAAGACGGCCTATCCGACN NNNNN	Reverse transcription
Mme_F	CTCAAGCTTCTAACGATGTACGCTCG A	First low cycle PCR & quality check
Mme_R	GCGGCTGAAGACGGCCTATCC	First low cycle PCR & quality check
TSS_Col3_short	GCCGTCTTCAGCCGCCTCAAGCTTCT AACGATGTACG	Bridge oligo for circularization
PE2_AN2	ACCGAGATCTACACTCTTTCCCTACA CGACGCTCTTCCGATCTNN	Illumina adapterA
PE2_AN2b	/5Phos/AGATCGGAAGAGCGTCGTGTA GGGAAAGAGTGTAGATCTCGGT	Illumina adapterA
PE2_BN2	/5Phos/AGATCGGAAGAGCGGTTCAG CAGGAATGCCGAGACCGATCT	Illumina adapterB
PE2_BN2b	AGATCGGTCTCGGCATTCCTGCTGAA CCGCTCTTCCGATCTNN	Illumina adapterB
PE2_AFN2_short	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACA	Final pcr
PE2_BRN2_short	CAAGCAGAAGACGGCATACGAGATC GGTCTCGGCATTCCT	Final pcr
PE2_A_short	AATGATACGGCGACCACCGAGA	Final pcr
PE2_B_short	CAAGCAGAAGACGGCATACGAGA	Final pcr

Table AI. 1: Oligos used for constructing TSS mapping libraries.

Nucleosomal DNA Isolation:

Formaldehyde-fixed yeast cultures were pelleted and washed. Cell pellets were resusupended in 39mL Buffer Z (1M Sorbital, 50mM Tris pH7.4) containing 28µL β-mercapoethanol. Cells were spheroblasted in this buffer at 30°C for approximately 40 minutes. The spheroblasts were pelleted at ~3500rcf for 10 minutes at 4°C. The cells were resuspended up to about 2.4mL with NP buffer

(50mM sodium chloride, 10mM Tris pH7.4, 5mM magnesium chloride, 1mM calcium chloride), containing 0.5mM spermidine, 1μL/mL β-mercaptoethanol, and 0.01% NP-40, and 0.6mL of cells was aliquoted to four tubes with micrococcal nuclease aliquoted in the lid. The cells were mixed with the enzyme and incubated at 37°C for 20 minutes, after which 150μL of STOP buffer (50% sodium dodecylsulfate, 0.05M ethylenediaminetetraacetic acid) and 5μL 20mg/mL proteinase K and the digestions were incubated at 65°C overnight to remove the histone proteins. Mononucleosomal DNA was purified by a PCI extraction and precipitation with 0.3M sodium acetate in 2-propanol and then dissolved in 60μL of 1X NEB buffer 2. The DNA was digested with 2μL of 20mg/mL RNase solution (Sigma) at 37°C for 1 hour. Digestion ladders were assessed by running 5μL on a 2% agarose gel. Digestions were chosen to have mostly mononucleosomal-sized DNA with a little dinucleosomal and a hint of trinucleosomal DNA. Half (25μL) of the appropriate digestion was treated with calf intestinal phosphatase (0.75μL) at 37°C for 45 minutes. The mononucleosomal band was gel purified from a 1.8% agarose gel, using Freeze N' Squeeze columns, and was PCI extracted and ethanol precipitated overnight.

Nucleosomal Deep Sequencing Library Preparation:

About 250ng of mononucleosomal DNA was end-cleaned with END-it (Epicentre) in a 50μL reaction. Qiagen MinElute columns were used to purify the reaction and DNA was eluted in 20μL water. DNA was then A-tailed with Klenow exo-

DNA Polymerase (Epicentre) in a 50μL reaction. This reaction was cleaned up on Qiagen MinElute columns and elutes with 10μL water to which 0.5μL Illumina genomic adapters was added; a 15μL ligation reaction was performed at room temperature for 1 hour and the reaction mixture was made up to 25μL and the ligation was continued at 16°C overnight. The ligation reaction was cleaned up on Qiagen MinElute columns and eluted with 20μL water. This was amplified with Pfx PCR reaction (Invitrogen) containing 0.5μL each Illumina genomic primers 1.1 and 2.1, 0.6mM dNTPs, 1mM magnesium sulfate, and 1uL Pfx. A portion of the PCR reactions (4μL) was run on a gel to check the presence and size of the product. The final PCR product was purified with Freeze N' Squeeze (BioRad) from a 1.8% agarose gel. One tenth of the purified library was used to quality check the library by StrataCloning, which requires an initial A-tailing reaction with Taq DNA Polymerase. 30μL of approximately 10nM amplified libraries were submitted for GAIIX Illumina deep sequencing.

Data Analysis, Nucleosomes:

Deep sequencing reads were aligned to sacCer2 using blat, and only uniquely aligned reads with fewer than 3 mismatches were retained. The cross correlation between reads mapping to the Watson and Crick strands were determined and reads were extended by this amount to generate nucleosome fragments. Coverage was counted at each base pair and the genome wide average count was normalized to one. Counts were then aligned around the edge of the +1 nucleosome as defined in (Tsankov et al., 2010). In order to call nucleosome

centre peaks, the edge of reads were added up and used in a Template Filtering code developed in (Weiner et al., 2010). The +1 and -1 nucleosome were defined as the nucleosomes flanking a nucleosome depleted region (linker between edges of called nucleosomes greater than 100) upstream of a coding region.

Data Analysis, TSS:

Reads were filtered to keep only reads from the TSS side, which ended in the reverse complement of the chimeric linker sequence (ending in GTTGGACTCGAGCGTACATCGTTAGAAGCTTGAG), which was trimmed. These trimmed reads were aligned to sacCer3 with bowtie2 defaults, and uniquely (MAPQ>10) mapped reads were kept. The TSS represents the reverse complement of the final base pair, so SAM FLAG representing Crick mapping were retained as TSS for Watson genes and vice versa. TSS counts were added up for each base pair genome wide and these counts were aligned surrounding the edge of the +1 nucleosome as defined in (Tsankov et al., 2010). For observing TSS shifts, TSS peaks were defined as the location of the highest value in a 200bp window surrounding the edge of the +1 nucleosome.

CHAPTER AII: The involvement of the preinitiation complex in promoter nucleosome occupancy and turnover

Introduction

Nucleosome packaging has been recognized to be dynamic; stable binding of histones would prevent the inducible expression of genes that are repressed by nucleosome packaging. Nucleosome turnover, as measured by the replacement of histone octamers by inducibly expressed, tagged histones, is highest at promoters (Dion et al., 2007). In fact, promoter nucleosomes are “hot” regardless of the gene body nucleosome turnover. While nucleosome turnover in the gene body correlates with RNA Polymerase II occupancy, promoter nucleosome turnover does not (Dion et al., 2007). Nucleosome depletion at yeast promoters is generally established by both sequence and *trans* acting factors, as to some extent transcription factors compete with histones to generate nucleosome depletion (Workman and Kingston, 1992). The preinitiation complex itself might help to maintain nucleosome depletion at promoters. Dynamic competition between histones and transcription factors could conceivably lead to turnover in the promoter. In order to determine the role of the preinitiation complex in promoter turnover and NDR formation, we measured turnover in the absence of the TATA binding protein (TBP) in a TBP anchor away strain.

Results and Discussion

Depletion of TBP, and hence PIC formation, does not dramatically affect average promoter depletion (Fig. AII-1A). Therefore the PIC does not appear to be globally responsible for nucleosome depletion; as anticipated, other transcription factors or chromatin remodelers play upstream roles in NDR formation, prior to the binding of TBP. Additionally, sequence-mediated depletion could be involved in the depletion of promoters thus allowing for maintained nucleosome depletion in the absence of TBP. However, TATA-containing promoters do experience some nucleosome fill-in of their NDRs, as compared to the relatively invariant TATA-less promoters (Fig. AII-1B).

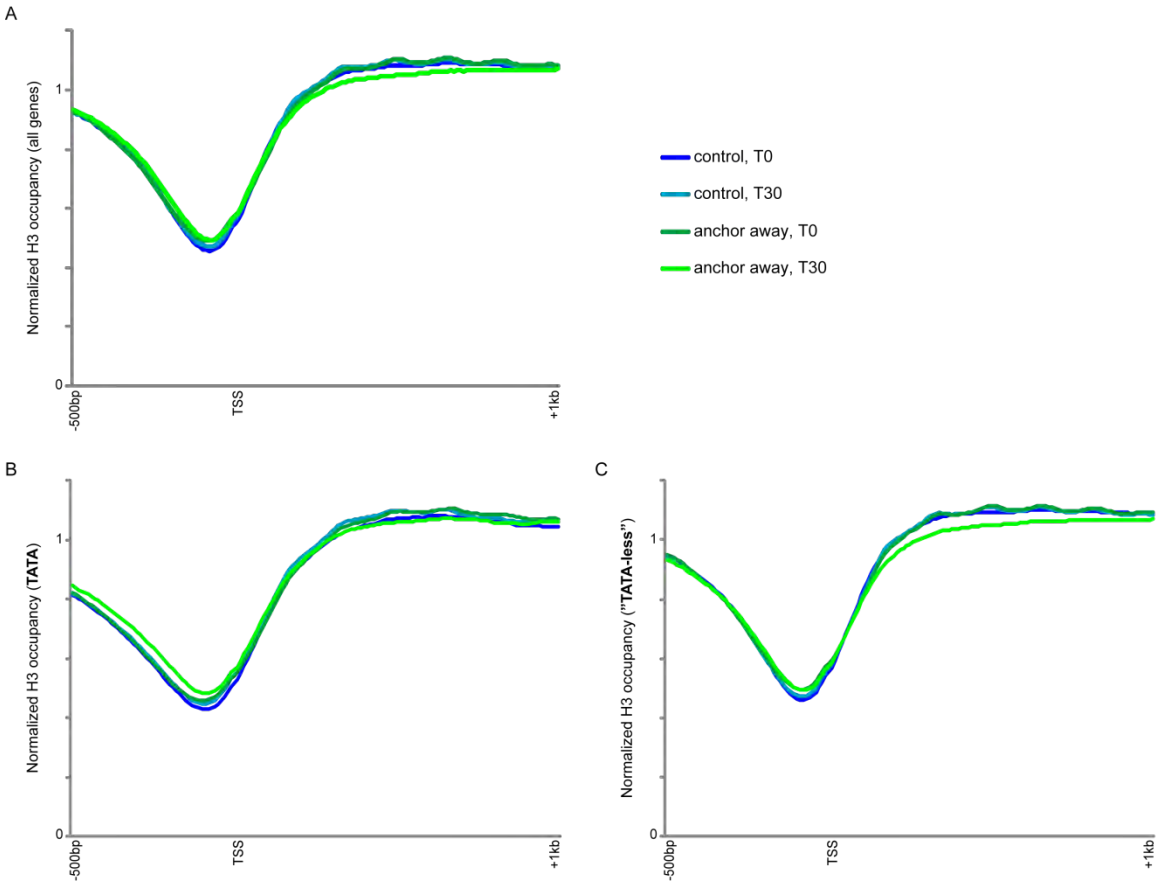


Figure AII. 1: TBP-depletion has a slight effect on promoter and genic nucleosome occupancy.

A) TSS alignments show that on average NDR formation is slightly decreased with the loss of TBP, as is genic nucleosome occupancy.
B) Average TSS alignments for TATA-containing genes shows a greater dependence on TBP for nucleosome depletion at these genes.
C) As (B) for TATA-less genes, showing that NDR formation is independent of TBP, but genic nucleosome occupancy is more greatly affected.

TATA-containing genes tend to be less stably expressed than TATA-less genes and their expression is more often dependent upon chromatin remodelers (Basehoar et al., 2004; Ioshikhes et al., 2006); they also display a less prominent NDR and less well-positioned nucleosome arrays (Albert et al., 2007; Tirosh and Barkai, 2008). It appears that the PIC may help to maintain or direct nucleosome depletion at these promoters, potentially via the recruitment of remodelers. It is also possible that these genes are induced with rapamycin treatment and hence experience transcription-associated nucleosome depletion in the TBP anchor away strain at 30 minutes; however the anchor away strain background harbors mutations that prevent rapamycin-induced expression changes, and there was no detected increase in gene expression of selected ribosomal protein genes (data not shown).

Nucleosome turnover in the control, untreated with rapamycin, is highest on average at the -1 nucleosome upstream of the NDR (Fig. AII-2A, centre panel); this holds true during the rapamycin induced depletion of TBP (Fig. AII-2A, left panel). In general, nucleosome turnover is slightly reduced in the TBP depleted conditions (Fig.AII- 2A, right panel; Fig. AII-2B), this is particularly true for TATA-less genes, where newly incorporated HA is not increased over the background level as compared to the control. These promoters also tend to experience less turnover, so TBP does not appear to be a major regulator of promoter nucleosome turnover. Much like promoter nucleosome depletion, it appears that

other transcription factors or chromatin remodelers define the high turnover rates seen at yeast promoters. The greater effect seen at TATA-less promoters might reflect the lesser role of other chromatin regulators in gene expression and promoter nucleosome architecture seen at these genes (Basehoar et al., 2004; Tirosh and Barkai, 2008).

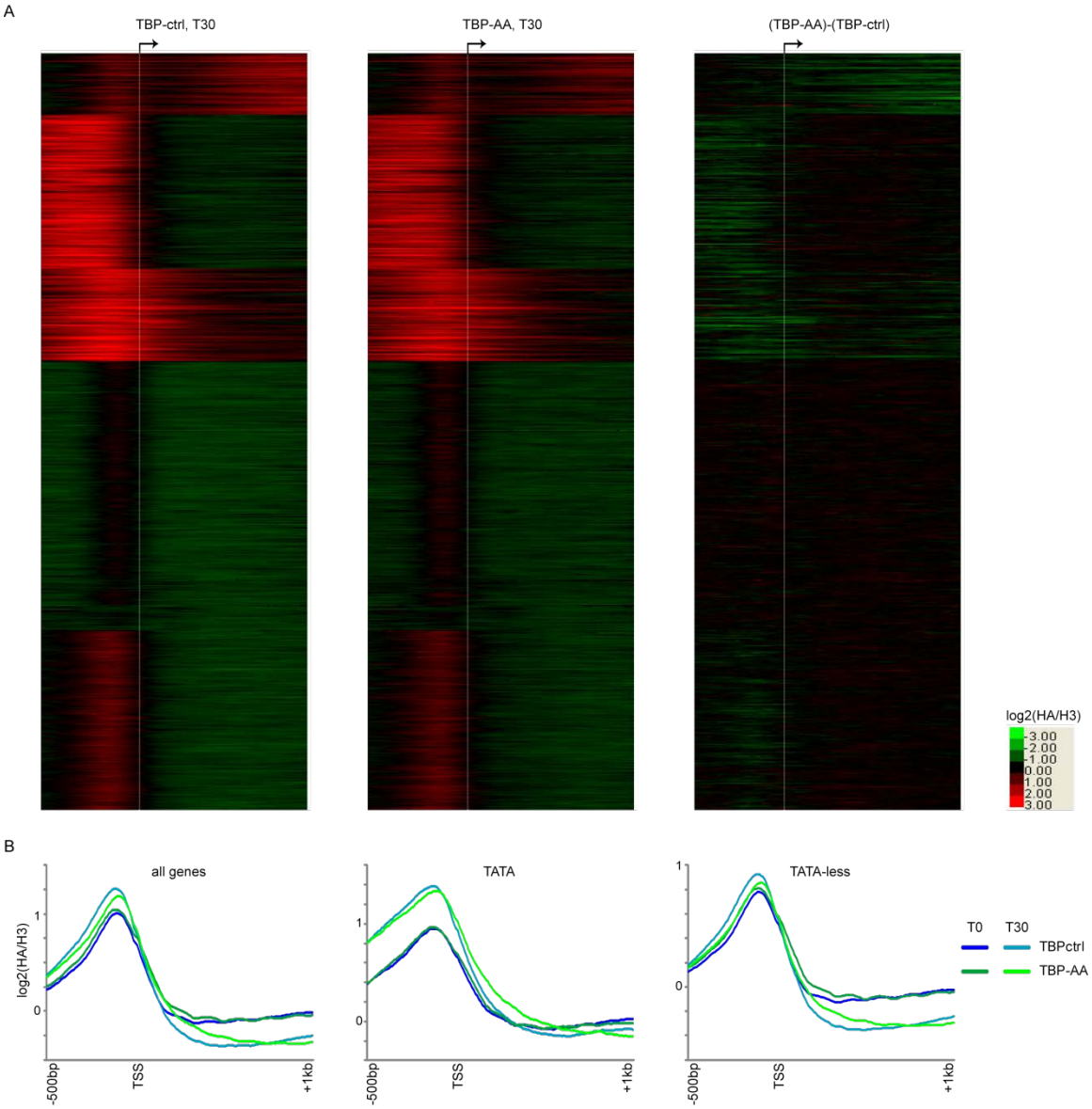


Figure AII. 2: Loss of TBP slightly decreases histone turnover.

A) Heatmaps showing the log2 of the enrichment of H3-HA over total H3 at 30 minutes after TBP depletion in an anchor away strain (centre panel) or in a control (left panel), along with the difference map (right panel), which shows general decreae in histone turnover, but conservation of histone turnover patterns.

B) Average histone turnover profiles surrounding the TSS shown for all (left), TATA (centre), and TATA-less (right) genes shows that promoter histone turnover is most affected by TBP at TATA-containing genes.

Materials and Method

Yeast strains and growth:

A C-terminal triple HA-tagged H3 under the control of a galactose inducible promoter on a plasmid with an ADE2 selectable marker was used for ChIP analysis of histone turnover. These strains also contained an *SPT6-FRB* generated by the integration of an *SPT6::KANMX6* pcr product at the C-terminus of *SPT6*. Strains were generated in the anchor away background strain, HHY221 (relevant genotype *MAT**a***, *tor1-1*, *fpr1::loxP-LEU2-loxP*, *RPL13A-2×FKBP12::loxP*, *ade2-1*, *trp1-1*, *his3-11*, *ura3*). Additionally *BAR1* was disrupted with a *loxP-K/URA3-loxP* cassette.

Yeast were grown to an OD_{600} of ~0.4-0.5 in Casamino acid medium lacking adenine with 2% raffinose and 0.1% glucose, at which point they were arrested in G1 with 600ng/mL of alpha factor (Primm srl, Italy) for 3 hours. HA-tagged histone expression was induced with the addition of 2% galactose, and concurrrently to induce depletion of TBP from the nucleus by anchor away rapamycin (LC laboratories) was added at a final concentration of 4µg/mL. While for controls 0.5% glucose was added instead of rapamycin to reduce the galactose induction of HA-tagged H3. For chromatin immunoprecipitation (ChIP), 100mL aliquots of culture were removed just prior to these additions (T0) and at 30 minutes post induction of HA-tagged H3 and TBP depletion.

Chromatin Immunoprecipitation:

Cross-linking was performed with 1.2% formaldehyde at 30°C for 10 minutes and quenched with 330mM glycine for 5 minutes at room temperature. Cells were washed with cold Tris buffered saline and resuspended in FA lysis buffer (50mM HEPES-KOH, pH 7.5, 140mM sodium chloride, 1mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate) with 1mM PMSF. Cells were broken with an equal volume of acid-washed glass beads with 40 minutes of vortexing at 4°C. Cells were pelleted and resuspended with 0.5% SDS. A Sonifier Cell Disruptor B-30 with duty 90 and output limit 250 on pulsing 10 times with 1 minute on ice in between was used to shear chromatin. Cell debris were pelleted, and the supernatant was sonicate 3 times. 30uL protein A-Sepharose beads, which had been preincubated with 2µg/mL herring sperm DNA and 20µg/mL BSA in FA-lysis buffer and washed twice with FA-lysis buffer, was combined with 100µL extract, 800µL FA-lysis buffer, and antibody at 4°C overnight. HA-tagged H3 and total H3 were immunoprecipitated with anti-HA antibodies (2µL, clone 16B12, Covance) and anti-H3 antibodies (1µL, #1791, Abcam), respectively. Beads were washed once each with 1mL FA-lysis buffer, FA500-lysis buffer (50mM HEPES-KOH, pH 7.5, 500mM sodium chloride, 1% Triton X-100, 0.1% sodium deoxycholate), Buffer III (10mM Tris-HCl, pH 8.0, 1mM EDTA, 250mM lithium chloride, 1% NP-40, 1% sodium deoxycholate), and twice with Tris-EDTA pH 8. The ChIP material was precipitated twice from the beads with 100µL elution buffer B (50mM Tris-HCl, pH 7.5, 1% sodium dodecyl sulfate, 10mM EDTA) at

60°C for 10 minutes and the eluants were combined. Crosslinks were reversed with the addition of 200µL Tris-EDTA pH 8 and 3µL proteinase K (20mg/mL) for 4 hours to overnight at 65°C. DNA was extract twice with phenol chloroform and once with chloroform followed by ethanol precipitation. Precipitated DNA was washed with cold 70% ethanol and dissolved in 100µL water.

Deep Sequencing Library Construction:

Samples were treated with RNaseA solution (Sigma, 20mg/mL) for 1 hour at 37°C in 1X NEB buffer 2 followed by a 45 minuted Calf Intestinal Phosphatase treatment. DNA was extracted with phenol chloroform isoamyl alcohol and ethanol precipitated with GlycoBlue. DNA ends were cleaned up with END-it (Epicentre) at room temperature for 1 hour. AMPure XP beads (Beckman-Coultier) were used to perform a double size selection, first depleting longer fragments with 0.5X beads and applying the supernatant to 1.3X more beads to size select upwards of 100bp. The beads were resuspended and A-tailing was performed on the beads with Klenow exo- polymerase (Epicentre). Cleanup was performed on the beads with the addition of 1.8x ABR buffer (15%PEG, 2.5M sodium chloride) and two 70% ethanol washes. The beads were resuspended and DNA was ligated with NEXTFlex (Bioo Scientific) multiplexed adapters on the beads (FastLink, Epicentre). Cleanup was performed as above, expect DNA was eluted with 39µL water and 2/3 of this was used for 18 and 20 cycle Pfx (Invitrogen) PCR reactions; the better of the two reaction conditions (as

determined by visualization on an agarose gel) was repeated with the final 1/3 of material. StrataClone was performed on 1/10 of the final PCR product to test the quality of the library. Samples were mixed in equimolar amounts (determined by KAPA Library Quantitation) and submitted to the UMass Deep Sequencing Core for HiSeq 100SR with multiplex indexing.

Data Analysis:

Sequence data was converted to a fasta format and aligned to the sacCer3 genome via blat. Only uniquely mapped reads with fewer than 3 mismatches were retained for mapping nucleosomes. In order to generate nucleosome profiles the cross correlation of the reads mapping to the Watson and Crick strands was determined and reads were extended by this value. The number of reads per base pair was counted and the genome wide average was normalized to one. Normalized counts were aligned surrounding the edge of the +1 nucleosome (as defined in (Tsankov et al., 2010)) and were averaged in 10bp bins. In order to observe enrichment for newly incorporated, HA-tagged histones, the log2 value of the ration of HA to H3 was used.

CHAPTER AIII: Mix n' matching whole cell extracts for *in vitro* reconstitution of chromatin structure

Introduction

Nucleosome positioning is influenced by both *cis* and *trans* acting factors. While the sequence content of *S.cerevisiae* promoters helps to deplete the region of nucleosomes (Kaplan et al., 2009), *trans* acting factors play a large role in dictating nucleosome positions. This was clearly demonstrated by the *in vitro* reconstitution of chromatin in yeast whole cell extract; here ATPase activity of the whole cell extract was required for the reconstitution of more native-like chromatin structure (Zhang et al., 2011). Additionally, in Chapter III we have validated the importance of *trans* acting factors to nucleosome depletion and positioning in an *in vivo* context (Hughes et al., 2012). The whole cell extract reconstitution system presents an elegant method that could be used to determine the factors necessary and sufficient for *in vivo* nucleosome architecture. This system has in fact been used with further purification of the whole cell extract to pinpoint the RSC chromatin remodeling complex as necessary for chromatin organization at some promoters (Wippo et al., 2011). In order to determine if whole cell extracts from a yeast species could reconstitute chromatin structure on heterologous DNA sequence, similarly to what we observed with yeast artificial chromosomes *in vivo*, we mapped nucleosomes on

in vitro reconstituted chromatin in the presence of whole cell extract from different yeast species.

Results and Discussion

We mapped nucleosomes on *S.cerevisiae*, *K.lactis*, and *S.pombe* sequence which had been reconstituted in the presence of ATP-supplemented whole cell extract from each of these species. In all cases, the matched sequence and whole cell extract performed the best in generating nucleosome depleted promoters flanked by well-positioned nucleosomes, however this is still incapable of forming *in vivo* nucleosomal arrays (Fig. AIII-1). The inability of species' whole cell extracts to substitute for one another may represent the specialization of *trans* factors for the sequence content of the corresponding genome. While NDRs are partially formed on *S.cerevisiae* and *K.lactis* sequence by salt gradient dialysis alone, whole cell extract from both of these species helps to further deplete these regions. This suggests that the sequence in these species' NDRs is hardwired for some depletion, but *trans* acting factors that are present in their whole cell extracts are also involved in depletion. It appears that the *S.cerevisiae* extract contains more abundant or more potent *trans* acting factors involved in nucleosome depletion, as the *S.cerevisiae* extract has greater nucleosome depletion activity on *K.lactis* sequence than the *K.lactis* extract. Interestingly, *S.pombe* extract produced the poorest nucleosome profiles, perhaps consistent with its divergence from budding yeast. Furthermore, *S.pombe* extract did not

lead to greater nucleosome depletion than salt gradient dialysis alone, indicating that *trans* acting factors involved in *S.cerevisiae* and *K.lactis* NDR formation are not conserved in *S.pombe.* However given the lack of well-positioned nucleosomes in the reconstitution of *S.pombe* sequence with its own extract, this may represent degradation of the extract or the presence of a nuclease in *S.pombe* extracts that has been thought to plague *in vivo* MNase sequencing in this species (Lantermann et al., 2010).

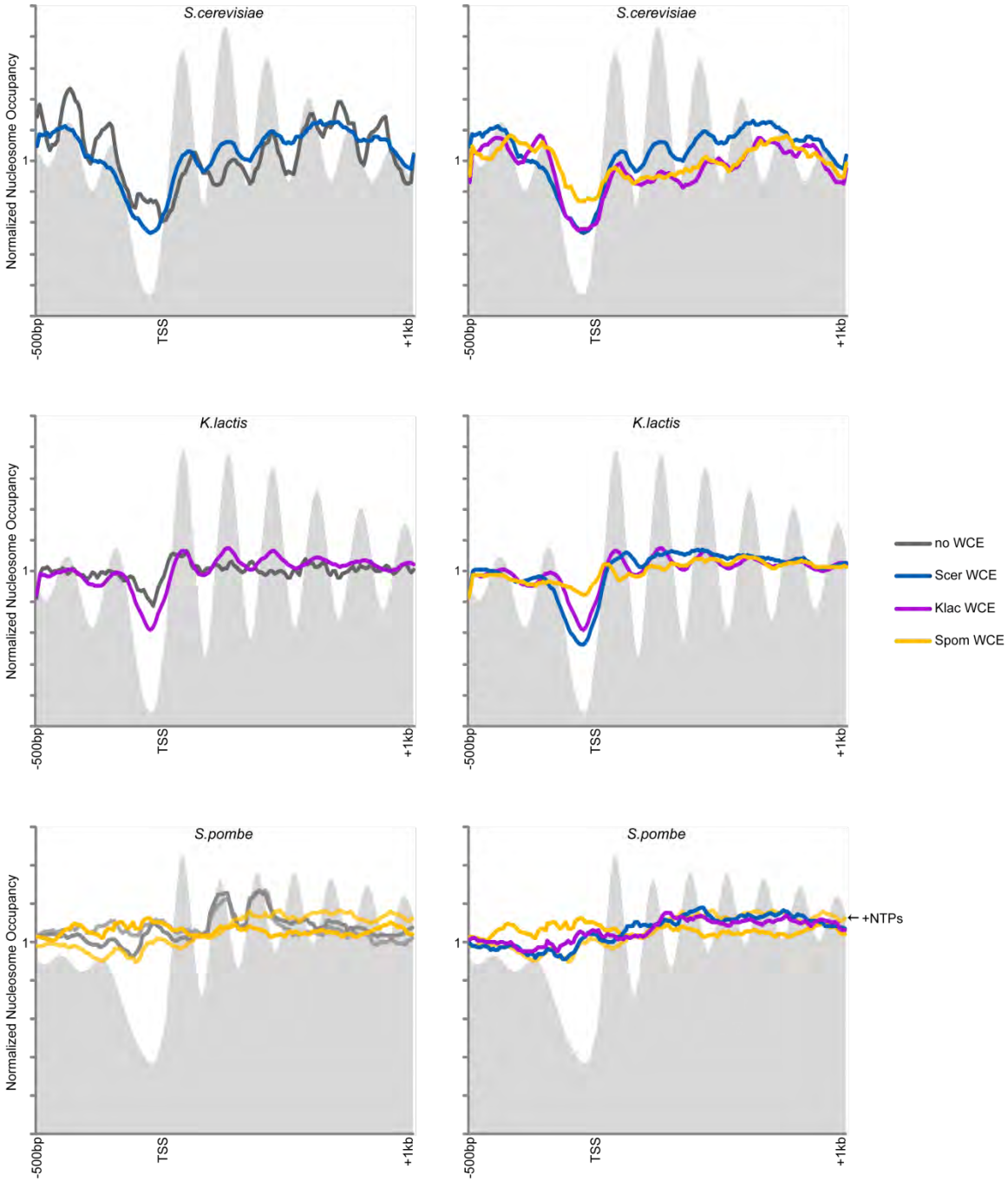

S.cerevisiae
S.cerevisiae
K.lactis
K.lactis
S.pombe
S.pombe
Normalized Nucleosome Occupancy
1
-500bp
TSS
+1kb
no WCE
Scer WCE
Klac WCE
Spom WCE
+NTPs

Figure AIII. 1: Reconstitution with WCE from heterologous species does not generate *in vivo* nucleosome profiles.

Reconstituted chromatin with *S.cerevisiae* (blue), *K.lactis* (purple), *S.pombe* (yellow), or no (grey) whole cell extract on *S.cerevisiae* (top), *K.lactis* (middle), and *S.pombe* (bottom) DNA is overlayed on *in vivo* TSS alignment averages for these species, shown in area fill. *S.pombe* extract performs poorly in all reconstitutions, while the other two extracts improve nucleosome depletion as well as some nucleosome positioning on *S.cerevisiae* and *K.lactis* DNA over salt gradient dialysis alone. In all cases, *in vivo* nucleosome maps show the most well-positioned nucleosomes.

Nucleosome depletion was the most conserved feature in general, as is consistent with the enrichment of polyA tracts in yeast promoters (Field et al., 2008; Kaplan et al., 2009). Interesting, unlike in the *in vivo* YAC context, the *S.cerevisiae* whole cell extract was unable to generate *S.cerevisiae* spacing of nucleosomes on *K.lactis* sequence, and nor did the application of *K.lactis* whole cell extract increase internucleosomal distance, as we might expect from the *trans* regulation of nucleosome spacing observed with YACs in Chapter III (Hughes et al., 2012). Nucleosomes are less well-positioned by whole cell extract than they are in the *in vivo* context, making interdyad measurements difficult. This could indicate that either whatever factor or factors are responsible for the measurement difference are unstable or depleted in the whole cell extract, or it could be consistent with a role for elongating polymerase in recruiting spacing factors. This latter explanation would fit with our observation in Chapter V that *K.lactis* Chd1 is able to generate wider nucleosome spacing in *S.cerevisiae*, particularly at genes with greater RNA Polymerase II abundance.

Materials and Methods

Reconstitutions and nucleosomal DNA isolation were performed by Nils Krietenstein in the Korber lab as described in (Zhang et al., 2011).

Nucleosomal library construction was performed by Megha Wal in the Pugh lab as described in (Zhang et al., 2011), and was sequenced on Illumina HiSeq. The 45bp reads were mapped to the relevant genome.

Data Analysis:

Watson and Crick mapped reads were counted up genome wide using bedtools genome coverage (Spom sequence) or a homemade perl script to count overlapping fragments, and the cross correlation between these reads was determined to infer the nucleosomal fragment length. Reads were extended to represent the nucleosome fragment and were recounted as above. The genome-wide average was normalized to one and these counts were aligned surround the edge of the +1 nucleosome, as defined in (Tsankov et al., 2010). These maps of reconstituted chromatin were compared to previously sequenced nucleosome maps of *S.cerevisiae* ((Zhang et al., 2011), “in vivo”), *K.lactis* (Tsankov et al., 2010), and *S.pombe* (Tsankov et al., 2011).

Bibliography

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

Allan, J., Fraser, R.M., Owen-Hughes, T., and Keszenman-Pereyra, D. (2012). Micrococcal nuclease does not substantially bias nucleosome mapping. J. Mol. Biol. *417*, 152–164.

Almer, A., and Horz, W. (1986). Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the PHO5/PHO3 locus in yeast. EMBO J. *5*, 2681–2687.

Almer, A., Rudolph, H., Hinnen, A., and Horz, W. (1986). Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. EMBO J. *5*, 2689–2696.

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

Anderson, J.D., and Widom, J. (2001). Poly(dA-dT) Promoter Elements Increase the Equilibrium Accessibility of Nucleosomal DNA Target Sites. Mol. Cell. Biol. *21*, 3830–3839.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L., et al. (2008). A Library of Yeast Transcription Factor Motifs Reveals a Widespread Function for Rsc3 in Targeting Nucleosome Exclusion at Promoters. Mol. Cell *32*, 878–887.

Bai, L., Charvin, G., Siggia, E.D., and Cross, F.R. (2010). Nucleosome-Depleted Regions in Cell-Cycle-Regulated Promoters Ensure Reliable Gene Expression in Every Cell Cycle. Dev. Cell *18*, 544–555.

Bai, L., Ondracka, A., and Cross, F.R. (2011). Multiple sequence-specific factors generate the nucleosome-depleted region on CLN2 promoter. Mol. Cell *42*, 465–476.

Van Bakel, H., Tsui, K., Gebbia, M., Mnaimneh, S., Hughes, T.R., and Nislow, C. (2013). A compendium of nucleosome and transcript profiles reveals determinants of chromatin architecture and transcription. PLoS Genet. *9*, e1003479.

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

Bell, O., Schwaiger, M., Oakeley, E.J., Lienert, F., Beisel, C., Stadler, M.B., and Schübeler, D. (2010). Accessibility of the Drosophila genome discriminates PcG repression, H4K16 acetylation and replication timing. Nat. Struct. Mol. Biol. *17*, 894–900.

Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. (2004). Global nucleosome occupancy in yeast. Genome Biol. *5*, R62.

Bintu, L., Kopaczynska, M., Hodges, C., Lubkowska, L., Kashlev, M., and Bustamante, C. (2011). The elongation rate of RNA polymerase determines the fate of transcribed nucleosomes. Nat. Struct. Mol. Biol. *18*, 1394–1399.

Biswas, D., Dutta-Biswas, R., and Stillman, D.J. (2007). Chd1 and yFACT act in opposition in regulating transcription. Mol. Cell. Biol. *27*, 6279–6287.

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

Bondarenko, V.A., Steele, L.M., Ujvári, A., Gaykalova, D.A., Kulaeva, O.I., Polikanov, Y.S., Luse, D.S., and Studitsky, V.M. (2006). Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. Mol. Cell *24*, 469–479.

Bouazoune, K., Miranda, T.B., Jones, P.A., and Kingston, R.E. (2009). Analysis of individual remodeled nucleosomes reveals decreased histone-DNA contacts created by hSWI/SNF. Nucleic Acids Res. *37*, 5279–5294.

Brem, R.B., Yvert, G., Clinton, R., and Kruglyak, L. (2002). Genetic dissection of transcriptional regulation in budding yeast. Science *296*, 752–755.

Brogaard, K., Xi, L., Wang, J.-P., and Widom, J. (2012). A map of nucleosome positions in yeast at base-pair resolution. Nature *486*, 496–501.

Bryant, G.O., Prabhu, V., Floer, M., Wang, X., Spagna, D., Schreiber, D., and Ptashne, M. (2008). Activator control of nucleosome occupancy in activation and repression of transcription. PLoS Biol. *6*, 2928–2939.

Buchman, A.R., and Kornberg, R.D. (1990). A yeast ARS-binding protein activates transcription synergistically in combination with other weak activating factors. Mol. Cell. Biol. *10*, 887–897.

Carone, B.R., Hung, J.-H., Hainer, S.J., Chou, M.-T., Carone, D.M., Weng, Z., Fazzio, T.G., and Rando, O.J. (2014). High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. Dev. Cell *30*, 11–22.

Celona, B., Weiner, A., Di Felice, F., Mancuso, F.M., Cesarini, E., Rossi, R.L., Gregory, L., Baban, D., Rossetti, G., Grianti, P., et al. (2011). Substantial histone reduction modulates genomewide nucleosomal occupancy and global transcriptional output. PLoS Biol. *9*, e1001086.

Chasman, D.I., Lue, N.F., Buchman, A.R., LaPointe, J.W., Lorch, Y., and Kornberg, R.D. (1990). A yeast protein that influences the chromatin structure of UASG and functions as a powerful auxiliary gene activator. Genes Dev. *4*, 503–514.

Chen, W., and Struhl, K. (1985). Yeast mRNA initiation sites are determined primarily by specific sequences, not by the distance from the TATA element. EMBO J. *4*, 3273–3280.

Choi, J.K., and Kim, Y.-J. (2009). Intrinsic variability of gene expression encoded in nucleosome positioning sequences. Nat. Genet. *41*, 498–503.

Choi, J.K., Hwang, S., and Kim, Y.-J. (2008). Stochastic and regulatory role of chromatin silencing in genomic response to environmental changes. PloS One *3*, e3002.

Chua, E.Y.D., Vasudevan, D., Davey, G.E., Wu, B., and Davey, C.A. (2012). The mechanics behind DNA sequence-dependent properties of the nucleosome. Nucleic Acids Res. *40*, 6338–6352.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu. Rev. Biochem. *78*, 273–304.

Conant, G.C., and Wolfe, K.H. (2007). Increased glycolytic flux as an outcome of whole-genome duplication in yeast. Mol. Syst. Biol. *3*, 129.

Dalal, Y., Wang, H., Lindsay, S., and Henikoff, S. (2007). Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. PLoS Biol. *5*, e218.

Deaton, A.M., Webb, S., Kerr, A.R.W., Illingworth, R.S., Guy, J., Andrews, R., and Bird, A. (2011). Cell type-specific DNA methylation at intragenic CpG islands in the immune system. Genome Res. *21*, 1074–1086.

Dion, M.F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., and Rando, O.J. (2007). Dynamics of replication-independent histone turnover in budding yeast. Science *315*, 1405–1408.

Drew, H.R., and Travers, A.A. (1985). DNA bending and its relation to nucleosome positioning. J. Mol. Biol. *186*, 773–790.

Fan, X., Moqtaderi, Z., Jin, Y., Zhang, Y., Liu, X.S., and Struhl, K. (2010). Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3'-end formation. Proc. Natl. Acad. Sci. *107*, 17945–17950.

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

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

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

Fu, Y., Sinha, M., Peterson, C.L., and Weng, Z. (2008). The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. PLoS Genet. *4*, e1000138.

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

Gaffney, D.J., McVicker, G., Pai, A.A., Fondufe-Mittendorf, Y.N., Lewellen, N., Michelini, K., Widom, J., Gilad, Y., and Pritchard, J.K. (2012). Controls of nucleosome positioning in the human genome. PLoS Genet. *8*, e1003036.

Ganapathi, M., Palumbo, M.J., Ansari, S.A., He, Q., Tsui, K., Nislow, C., and Morse, R.H. (2011). Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast. Nucleic Acids Res. *39*, 2032–2044.

Giardina, C., and Lis, J.T. (1993). DNA melting on yeast RNA polymerase II promoters. Science *261*, 759–762.

Gilchrist, D.A., Dos Santos, G., Fargo, D.C., Xie, B., Gao, Y., Li, L., and Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. Cell *143*, 540–551.

Gkikopoulos, T., Schofield, P., Singh, V., Pinskaya, M., Mellor, J., Smolle, M., Workman, J.L., Barton, G.J., and Owen-Hughes, T. (2011). A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. Science *333*, 1758–1760.

Gossett, A.J., and Lieb, J.D. (2012). In vivo effects of histone H3 depletion on nucleosome occupancy and position in Saccharomyces cerevisiae. PLoS Genet. *8*, e1002771.

Gracey, L.E., Chen, Z.-Y., Maniar, J.M., Valouev, A., Sidow, A., Kay, M.A., and Fire, A.Z. (2010). An in vitro-identified high-affinity nucleosome-positioning signal is capable of transiently positioning a nucleosome in vivo. Epigenetics Chromatin *3*, 13.

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

Han, M., Kim, U.J., Kayne, P., and Grunstein, M. (1988a). Depletion of histone H4 and nucleosomes activates the PHO5 gene in Saccharomyces cerevisiae. EMBO J. *7*, 2221–2228.

Han, M., Kim, U.J., Kayne, P., and Grunstein, M. (1988b). Depletion of histone H4 and nucleosomes activates the PHO5 gene in Saccharomyces cerevisiae. EMBO J. *7*, 2221–2228.

Hao, N., and O'Shea, E.K. (2012). Signal-dependent dynamics of transcription factor translocation controls gene expression. Nat. Struct. Mol. Biol. *19*, 31–39.

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

Hauk, G., McKnight, J.N., Nodelman, I.M., and Bowman, G.D. (2010). The chromodomains of the Chd1 chromatin remodeler regulate DNA access to the ATPase motor. Mol. Cell *39*, 711–723.

He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M., et al. (2010). Nucleosome dynamics define transcriptional enhancers. Nat. Genet. *42*, 343–347.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M., and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. Proc. Natl. Acad. Sci. U. S. A. *108*, 18318–18323.

Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., et al. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat. Methods *6*, 283–289.

Heus, J.J., Zonneveld, B.J., Bloom, K.S., de Steensma, H.Y., and van den Berg, J.A. (1993). The nucleosome repeat length of Kluyveromyces lactis is 16 bp longer than that of Saccharomyces cerevisiae. Nucleic Acids Res. *21*, 2247–2248.

Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., and Bustamante, C. (2009). Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. Science *325*, 626–628.

Hogan, G.J., Lee, C.-K., and Lieb, J.D. (2006). Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. PLoS Genet. *2*, e158.

Van Holde, K. (1989). Chromatin. N. Y. Springer-Verl. 497.

Hornung, G., Oren, M., and Barkai, N. (2012). Nucleosome organization affects the sensitivity of gene expression to promoter mutations. Mol. Cell *46*, 362–368.

Hörz, W., and Altenburger, W. (1981). Sequence specific cleavage of DNA by micrococcal nuclease. Nucleic Acids Res. *9*, 2643–2658.

Hota, S.K., Bhardwaj, S.K., Deindl, S., Lin, Y., Zhuang, X., and Bartholomew, B. (2013). Nucleosome mobilization by ISW2 requires the concerted action of the ATPase and SLIDE domains. Nat. Struct. Mol. Biol. *20*, 222–229.

Hughes, A.L., Jin, Y., Rando, O.J., and Struhl, K. (2012). A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. Mol. Cell *48*, 5–15.

Huisinga, K.L., and Pugh, B.F. (2004). A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. Mol. Cell *13*, 573–585.

Ihmels, J., Bergmann, S., Gerami-Nejad, M., Yanai, I., McClellan, M., Berman, J., and Barkai, N. (2005). Rewiring of the yeast transcriptional network through the evolution of motif usage. Science *309*, 938–940.

Ioshikhes, I., Bolshoy, A., Derenshteyn, K., Borodovsky, M., and Trifonov, E.N. (1996). Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. J. Mol. Biol. *262*, 129–139.

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

Iyer, V., and Struhl, K. (1995). Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. EMBO J. *14*, 2570–2579.

Jessen, W.J., Hoose, S.A., Kilgore, J.A., and Kladde, M.P. (2006). Active PHO5 chromatin encompasses variable numbers of nucleosomes at individual promoters. Nat. Struct. Mol. Biol. *13*, 256–263.

Jin, J., Bai, L., Johnson, D.S., Fulbright, R.M., Kireeva, M.L., Kashlev, M., and Wang, M.D. (2010). Synergistic action of RNA polymerases in overcoming the nucleosomal barrier. Nat. Struct. Mol. Biol. *17*, 745–752.

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

Keene, M.A., and Elgin, S.C. (1981). Micrococcal nuclease as a probe of DNA sequence organization and chromatin structure. Cell *27*, 57–64.

Kelly, T.K., Liu, Y., Lay, F.D., Liang, G., Berman, B.P., and Jones, P.A. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res. *22*, 2497–2506.

Kim, T.S., Liu, C.L., Yassour, M., Holik, J., Friedman, N., Buratowski, S., and Rando, O.J. (2010). RNA polymerase mapping during stress responses reveals widespread nonproductive transcription in yeast. Genome Biol. *11*, R75.

Korber, P., and Hörz, W. (2004). In vitro assembly of the characteristic chromatin organization at the yeast PHO5 promoter by a replication-independent extract system. J. Biol. Chem. *279*, 35113–35120.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285–294.

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

Kostrewa, D., Zeller, M.E., Armache, K.-J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. Nature *462*, 323–330.

Kulaeva, O.I., Gaykalova, D.A., Pestov, N.A., Golovastov, V.V., Vassylyev, D.G., Artsimovitch, I., and Studitsky, V.M. (2009). Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. Nat. Struct. Mol. Biol. *16*, 1272–1278.

Kulaeva, O.I., Hsieh, F.-K., and Studitsky, V.M. (2010). RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones. Proc. Natl. Acad. Sci. U. S. A. *107*, 11325–11330.

Lam, F.H., Steger, D.J., and O'Shea, E.K. (2008). Chromatin decouples promoter threshold from dynamic range. Nature *453*, 246–250.

Lantermann, A.B., Straub, T., Strålfors, A., Yuan, G.-C., Ekwall, K., and Korber, P. (2010). Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning

mechanisms distinct from those of Saccharomyces cerevisiae. Nat. Struct. Mol. Biol. *17*, 251–257.

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

Lee, K., Kim, S.C., Jung, I., Kim, K., Seo, J., Lee, H.-S., Bogu, G.K., Kim, D., Lee, S., Lee, B., et al. (2013). Genetic landscape of open chromatin in yeast. PLoS Genet. *9*, e1003229.

Lee, S.-I., Pe'er, D., Dudley, A.M., Church, G.M., and Koller, D. (2006). Identifying regulatory mechanisms using individual variation reveals key role for chromatin modification. Proc. Natl. Acad. Sci. U. S. A. *103*, 14062–14067.

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

LeRoy, G., Orphanides, G., Lane, W.S., and Reinberg, D. (1998). Requirement of RSF and FACT for transcription of chromatin templates in vitro. Science *282*, 1900–1904.

Li, B., Pattenden, S.G., Lee, D., Gutiérrez, J., Chen, J., Seidel, C., Gerton, J., and Workman, J.L. (2005). Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. Proc. Natl. Acad. Sci. U. S. A. *102*, 18385–18390.

Li, Y., Flanagan, P.M., Tschochner, H., and Kornberg, R.D. (1994). RNA polymerase II initiation factor interactions and transcription start site selection. Science *263*, 805–807.

Lidor Nili, E., Field, Y., Lubling, Y., Widom, J., Oren, M., and Segal, E. (2010). p53 binds preferentially to genomic regions with high DNA-encoded nucleosome occupancy. Genome Res. *20*, 1361–1368.

Lomvardas, S., and Thanos, D. (2002). Modifying gene expression programs by altering core promoter chromatin architecture. Cell *110*, 261–271.

Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. *276*, 19–42.

Lu, Q., Wallrath, L.L., and Elgin, S.C. (1995). The role of a positioned nucleosome at the Drosophila melanogaster hsp26 promoter. EMBO J. *14*, 4738–4746.

Ludwigsen, J., Klinker, H., and Mueller-Planitz, F. (2013). No need for a power stroke in ISWI-mediated nucleosome sliding. EMBO Rep. *14*, 1092–1097.

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

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinforma. Oxf. Engl. *27*, 1696–1697.

Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature *429*, 571–574.

Marx, K.A., Zhou, Y., and Kishawi, I.Q. (2006). Evidence for long poly(dA).poly(dT) tracts in D. discoideum DNA at high frequencies and their preferential avoidance of nucleosomal DNA core regions. J. Biomol. Struct. Dyn. *23*, 429–446.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., et al. (2008a). Nucleosome organization in the Drosophila genome. Nature *453*, 358–362.

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

McKnight, J.N., Jenkins, K.R., Nodelman, I.M., Escobar, T., and Bowman, G.D. (2011). Extranucleosomal DNA binding directs nucleosome sliding by Chd1. Mol. Cell. Biol. *31*, 4746–4759.

McPherson, C.E., Shim, E.Y., Friedman, D.S., and Zaret, K.S. (1993). An active tissue-specific enhancer and bound transcription factors existing in a precisely positioned nucleosomal array. Cell *75*, 387–398.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat. Genet. *37*, 1090–1097.

Möbius, W., and Gerland, U. (2010). Quantitative test of the barrier nucleosome model for statistical positioning of nucleosomes up- and downstream of transcription start sites. PLoS Comput. Biol. *6*.

Möbius, W., Osberg, B., Tsankov, A.M., Rando, O.J., and Gerland, U. (2013). Toward a unified physical model of nucleosome patterns flanking transcription start sites. Proc. Natl. Acad. Sci. U. S. A. *110*, 5719–5724.

Morozov, A.V., Fortney, K., Gaykalova, D.A., Studitsky, V.M., Widom, J., and Siggia, E.D. (2009). Using DNA mechanics to predict in vitro nucleosome positions and formation energies. Nucleic Acids Res. *37*, 4707–4722.

Moshkin, Y.M., Chalkley, G.E., Kan, T.W., Reddy, B.A., Ozgur, Z., van Ijcken, W.F.J., Dekkers, D.H.W., Demmers, J.A., Travers, A.A., and Verrijzer, C.P. (2012). Remodelers organize cellular chromatin by counteracting intrinsic histone-DNA sequence preferences in a class-specific manner. Mol. Cell. Biol. *32*, 675–688.

Nagy, P.L., Cleary, M.L., Brown, P.O., and Lieb, J.D. (2003). Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. Proc. Natl. Acad. Sci. U. S. A. *100*, 6364–6369.

Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840–846.

Ni, T., Corcoran, D.L., Rach, E.A., Song, S., Spana, E.P., Gao, Y., Ohler, U., and Zhu, J. (2010). A paired-end sequencing strategy to map the complex landscape of transcription initiation. Nat. Methods *7*, 521–527.

Nodelman, I.M., and Bowman, G.D. (2013). Nucleosome sliding by Chd1 does not require rigid coupling between DNA-binding and ATPase domains. EMBO Rep. *14*, 1098–1103.

Noll, M. (1974). Subunit structure of chromatin. Nature *251*, 249–251.

Olson, W.K., and Zhurkin, V.B. (2011). Working the kinks out of nucleosomal DNA. Curr. Opin. Struct. Biol. *21*, 348–357.

Ong, M.S., Richmond, T.J., and Davey, C.A. (2007). DNA stretching and extreme kinking in the nucleosome core. J. Mol. Biol. *368*, 1067–1074.

Orphanides, G., LeRoy, G., Chang, C.H., Luse, D.S., and Reinberg, D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. Cell *92*, 105–116.

Owen-Hughes, T., and Workman, J.L. (1996). Remodeling the chromatin structure of a nucleosome array by transcription factor-targeted trans-displacement of histones. EMBO J. *15*, 4702–4712.

Parnell, T.J., Huff, J.T., and Cairns, B.R. (2008). RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. EMBO J. *27*, 100–110.

Patel, A., McKnight, J.N., Genzor, P., and Bowman, G.D. (2011). Identification of residues in chromodomain helicase DNA-binding protein 1 (Chd1) required for coupling ATP hydrolysis to nucleosome sliding. J. Biol. Chem. *286*, 43984–43993.

Patel, A., Chakravarthy, S., Morrone, S., Nodelman, I.M., McKnight, J.N., and Bowman, G.D. (2013). Decoupling nucleosome recognition from DNA binding dramatically alters the properties of the Chd1 chromatin remodeler. Nucleic Acids Res. *41*, 1637–1648.

Perales, R., Zhang, L., and Bentley, D. (2011). Histone occupancy in vivo at the 601 nucleosome binding element is determined by transcriptional history. Mol. Cell. Biol. *31*, 3485–3496.

Peterson, C.L., and Tamkun, J.W. (1995). The SWI-SNF complex: a chromatin remodeling machine? Trends Biochem. Sci. *20*, 143–146.

Pinto, I., Wu, W.H., Na, J.G., and Hampsey, M. (1994). Characterization of sua7 mutations defines a domain of TFIIB involved in transcription start site selection in yeast. J. Biol. Chem. *269*, 30569–30573.

Pointner, J., Persson, J., Prasad, P., Norman-Axelsson, U., Strålfors, A., Khorosjutina, O., Krietenstein, N., Svensson, J.P., Ekwall, K., and Korber, P. (2012). CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in S. pombe. EMBO J. *31*, 4388–4403.

Polach, K.J., and Widom, J. (1995). Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. J. Mol. Biol. *254*, 130–149.

Polach, K.J., and Widom, J. (1996). A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. J. Mol. Biol. *258*, 800–812.

Puhl, H.L., and Behe, M.J. (1995). Poly(dA).poly(dT) forms very stable nucleosomes at higher temperatures. J. Mol. Biol. *245*, 559–567.

Puhl, H.L., Gudibande, S.R., and Behe, M.J. (1991). Poly[d(A.T)] and other synthetic polydeoxynucleotides containing oligoadenosine tracts form nucleosomes easily. J. Mol. Biol. *222*, 1149–1160.

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

Radman-Livaja, M., Ruben, G., Weiner, A., Friedman, N., Kamakaka, R., and Rando, O.J. (2011). Dynamics of Sir3 spreading in budding yeast: secondary recruitment sites and euchromatic localization. EMBO J. *30*, 1012–1026.

Radman-Livaja, M., Quan, T.K., Valenzuela, L., Armstrong, J.A., van Welsem, T., Kim, T., Lee, L.J., Buratowski, S., van Leeuwen, F., Rando, O.J., et al. (2012). A key role for Chd1 in histone H3 dynamics at the 3' ends of long genes in yeast. PLoS Genet. *8*, e1002811.

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

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell *138*, 114–128.

Rando, O.J. (2012). Combinatorial complexity in chromatin structure and function: revisiting the histone code. Curr. Opin. Genet. Dev. *22*, 148–155.

Raser, J.M., and O'Shea, E.K. (2004). Control of stochasticity in eukaryotic gene expression. Science *304*, 1811–1814.

Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., and Segal, E. (2012). Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nat. Genet. *44*, 743–750.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295–301.

Rhodes, D. (1979). Nucleosome cores reconstituted from poly (dA-dT) and the octamer of histones. Nucleic Acids Res. *6*, 1805–1816.

Rosin, D., Hornung, G., Tirosh, I., Gispan, A., and Barkai, N. (2012). Promoter nucleosome organization shapes the evolution of gene expression. PLoS Genet. *8*, e1002579.

Rufiange, A., Jacques, P.-E., Bhat, W., Robert, F., and Nourani, A. (2007). Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. Mol. Cell *27*, 393–405.

Sasaki, S., Mello, C.C., Shimada, A., Nakatani, Y., Hashimoto, S.-I., Ogawa, M., Matsushima, K., Gu, S.G., Kasahara, M., Ahsan, B., et al. (2009). Chromatin-associated periodicity in genetic variation downstream of transcriptional start sites. Science *323*, 401–404.

Satchwell, S.C., Drew, H.R., and Travers, A.A. (1986). Sequence periodicities in chicken nucleosome core DNA. J. Mol. Biol. *191*, 659–675.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887–898.

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. Nat. Struct. Mol. Biol. *16*, 990–995.

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

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

Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., and Segal, E. (2012). Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat. Biotechnol. *30*, 521–530.

Simic, R., Lindstrom, D.L., Tran, H.G., Roinick, K.L., Costa, P.J., Johnson, A.D., Hartzog, G.A., and Arndt, K.M. (2003). Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. EMBO J. *22*, 1846–1856.

Simpson, R.T., and Künzler, P. (1979). Cromatin and core particles formed from the inner histones and synthetic polydeoxyribonucleotides of defined sequence. Nucleic Acids Res. *6*, 1387–1415.

Simpson, R.T., and Stafford, D.W. (1983). Structural features of a phased nucleosome core particle. Proc. Natl. Acad. Sci. U. S. A. *80*, 51–55.

Small, E.C., Xi, L., Wang, J.-P., Widom, J., and Licht, J.D. (2014). Single-cell nucleosome mapping reveals the molecular basis of gene expression heterogeneity. Proc. Natl. Acad. Sci. U. S. A. *111*, E2462–E2471.

Smolle, M., Venkatesh, S., Gogol, M.M., Li, H., Zhang, Y., Florens, L., Washburn, M.P., and Workman, J.L. (2012). Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. Nat. Struct. Mol. Biol. *19*, 884–892.

Stalder, J., Groudine, M., Dodgson, J.B., Engel, J.D., and Weintraub, H. (1980). Hb switching in chickens. Cell *19*, 973–980.

Stockdale, C., Flaus, A., Ferreira, H., and Owen-Hughes, T. (2006). Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. J. Biol. Chem. *281*, 16279–16288.

Studitsky, V.M., Clark, D.J., and Felsenfeld, G. (1994). A histone octamer can step around a transcribing polymerase without leaving the template. Cell *76*, 371–382.

Studitsky, V.M., Kassavetis, G.A., Geiduschek, E.P., and Felsenfeld, G. (1997). Mechanism of transcription through the nucleosome by eukaryotic RNA polymerase. Science *278*, 1960–1963.

Stünkel, W., Kober, I., and Seifart, K.H. (1997). A nucleosome positioned in the distal promoter region activates transcription of the human U6 gene. Mol. Cell. Biol. *17*, 4397–4405.

Svaren, J., Schmitz, J., and Hörz, W. (1994). The transactivation domain of Pho4 is required for nucleosome disruption at the PHO5 promoter. EMBO J. *13*, 4856–4862.

Teif, V.B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.-P., Marth, C., Höfer, T., and Rippe, K. (2012). Genome-wide nucleosome positioning during embryonic stem cell development. Nat. Struct. Mol. Biol. *19*, 1185–1192.

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., and Guigó, R. (2009). Nucleosome positioning as a determinant of exon recognition. Nat. Struct. Mol. Biol. *16*, 996–1001.

Tillo, D., and Hughes, T.R. (2009). G+C content dominates intrinsic nucleosome occupancy. BMC Bioinformatics *10*, 442.

Tillo, D., Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Field, Y., Lieb, J.D., Widom, J., Segal, E., and Hughes, T.R. (2010). High nucleosome occupancy is encoded at human regulatory sequences. PloS One *5*, e9129.

Tims, H.S., Gurunathan, K., Levitus, M., and Widom, J. (2011). Dynamics of nucleosome invasion by DNA binding proteins. J. Mol. Biol. *411*, 430–448.

Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. Genome Res. *18*, 1084–1091.

Tirosh, I., Sigal, N., and Barkai, N. (2010). Divergence of nucleosome positioning between two closely related yeast species: genetic basis and functional consequences. Mol. Syst. Biol. *6*, 365.

Tolkunov, D., Zawadzki, K.A., Singer, C., Elfving, N., Morozov, A.V., and Broach, J.R. (2011). Chromatin remodelers clear nucleosomes from intrinsically unfavorable sites to establish nucleosome-depleted regions at promoters. Mol. Biol. Cell *22*, 2106–2118.

Tolstorukov, M.Y., Colasanti, A.V., McCandlish, D.M., Olson, W.K., and Zhurkin, V.B. (2007). A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. J. Mol. Biol. *371*, 725–738.

Tomar, R.S., Psathas, J.N., Zhang, H., Zhang, Z., and Reese, J.C. (2009). A novel mechanism of antagonism between ATP-dependent chromatin remodeling complexes regulates RNR3 expression. Mol. Cell. Biol. *29*, 3255–3265.

Tsankov, A., Yanagisawa, Y., Rhind, N., Regev, A., and Rando, O.J. (2011). Evolutionary divergence of intrinsic and trans-regulated nucleosome positioning sequences reveals plastic rules for chromatin organization. Genome Res. *21*, 1851–1862.

Tsankov, A.M., Thompson, D.A., Socha, A., Regev, A., and Rando, O.J. (2010). The role of nucleosome positioning in the evolution of gene regulation. PLoS Biol. *8*, e1000414.

Udugama, M., Sabri, A., and Bartholomew, B. (2011). The INO80 ATP-dependent chromatin remodeling complex is a nucleosome spacing factor. Mol. Cell. Biol. *31*, 662–673.

Vaillant, C., Palmeira, L., Chevereau, G., Audit, B., d' Aubenton-Carafa, Y., Thermes, C., and Arneodo, A. (2010). A novel strategy of transcription regulation by intragenic nucleosome ordering. Genome Res. *20*, 59–67.

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K., et al. (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res. *18*, 1051–1063.

Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. Nature *474*, 516–520.

Vary, J.C., Gangaraju, V.K., Qin, J., Landel, C.C., Kooperberg, C., Bartholomew, B., and Tsukiyama, T. (2003). Yeast Isw1p forms two separable complexes in vivo. Mol. Cell. Biol. *23*, 80–91.

Vavouri, T., and Lehner, B. (2012). Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. Genome Biol. *13*, R110.

Verzi, M.P., Shin, H., He, H.H., Sulahian, R., Meyer, C.A., Montgomery, R.K., Fleet, J.C., Brown, M., Liu, X.S., and Shivdasani, R.A. (2010). Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. Dev. Cell *19*, 713–726.

Vinces, M.D., Legendre, M., Caldara, M., Hagihara, M., and Verstrepen, K.J. (2009). Unstable tandem repeats in promoters confer transcriptional evolvability. Science *324*, 1213–1216.

Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y., et al. (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res. *22*, 1798–1812.

Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A. (2007). Natural history and evolutionary principles of gene duplication in fungi. Nature *449*, 54–61.

Weiner, A., Hughes, A., Yassour, M., Rando, O.J., and Friedman, N. (2010). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res. *20*, 90–100.

Weintraub, H., and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science *193*, 848–856.

West, S.M., Rohs, R., Mann, R.S., and Honig, B. (2010). Electrostatic interactions between arginines and the minor groove in the nucleosome. J. Biomol. Struct. Dyn. *27*, 861–866.

Westenberger, S.J., Cui, L., Dharia, N., Winzeler, E., and Cui, L. (2009). Genome-wide nucleosome mapping of Plasmodium falciparum reveals histone-rich coding and histone-poor intergenic regions and chromatin remodeling of core and subtelomeric genes. BMC Genomics *10*, 610.

Whitehouse, I., and Tsukiyama, T. (2006). Antagonistic forces that position nucleosomes in vivo. Nat. Struct. Mol. Biol. *13*, 633–640.

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

Wilhelm, B.T., Marguerat, S., Aligianni, S., Codlin, S., Watt, S., and Bähler, J. (2011). Differential patterns of intronic and exonic DNA regions with respect to RNA polymerase II occupancy, nucleosome density and H3K36me3 marking in fission yeast. Genome Biol. *12*, R82.

Winston, F., and Carlson, M. (1992). Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. Trends Genet. TIG *8*, 387–391.

Wippo, C.J., Israel, L., Watanabe, S., Hochheimer, A., Peterson, C.L., and Korber, P. (2011). The RSC chromatin remodelling enzyme has a unique role in directing the accurate positioning of nucleosomes. EMBO J. *30*, 1277–1288.

Workman, J.L., and Kingston, R.E. (1992). Nucleosome core displacement in vitro via a metastable transcription factor-nucleosome complex. Science *258*, 1780–1784.

Workman, J.L., Taylor, I.C., and Kingston, R.E. (1991). Activation domains of stably bound GAL4 derivatives alleviate repression of promoters by nucleosomes. Cell *64*, 533–544.

Wyrick, J.J., Holstege, F.C., Jennings, E.G., Causton, H.C., Shore, D., Grunstein, M., Lander, E.S., and Young, R.A. (1999). Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. Nature *402*, 418–421.

Xi, Y., Yao, J., Chen, R., Li, W., and He, X. (2011). Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. Genome Res. *21*, 718–724.

Yamada, K., Frouws, T.D., Angst, B., Fitzgerald, D.J., DeLuca, C., Schimmele, K., Sargent, D.F., and Richmond, T.J. (2011). Structure and mechanism of the chromatin remodelling factor ISW1a. Nature *472*, 448–453.

Yarragudi, A., Miyake, T., Li, R., and Morse, R.H. (2004). Comparison of ABF1 and RAP1 in chromatin opening and transactivator potentiation in the budding yeast Saccharomyces cerevisiae. Mol. Cell. Biol. *24*, 9152–9164.

Yen, K., Vinayachandran, V., Batta, K., Koerber, R.T., and Pugh, B.F. (2012). Genome-wide nucleosome specificity and directionality of chromatin remodelers. Cell *149*, 1461–1473.

Yu, L., and Morse, R.H. (1999). Chromatin opening and transactivator potentiation by RAP1 in Saccharomyces cerevisiae. Mol. Cell. Biol. *19*, 5279–5288.

Yuan, G.-C., and Liu, J.S. (2008). Genomic sequence is highly predictive of local nucleosome depletion. PLoS Comput. Biol. *4*, e13.

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

Zentner, G.E., Tsukiyama, T., and Henikoff, S. (2013). ISWI and CHD chromatin remodelers bind promoters but act in gene bodies. PLoS Genet. *9*, e1003317.

Zhang, Z., and Dietrich, F.S. (2005). Mapping of transcription start sites in Saccharomyces cerevisiae using 5’ SAGE. Nucleic Acids Res. *33*, 2838–2851.

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

Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., and Pugh, B.F. (2011). A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. Science *332*, 977–980.