The Dissertation Committee for Dia Nicholson Bagchi Certifies that this is the approved version of the following dissertation:

The Interplay between Transcription, Histone Variants, and Chromatin Structure in Eukaryotes

Committee:

Vishwanath R. Iyer, Supervisor

Edward M. Marcotte

Claus O. Wilke

Kyle M. Miller

Haley Tucker

The Interplay between Transcription, Histone Variants, and Chromatin Structure in Eukaryotes

by

Dia Nicholson Bagchi, B.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2016

Dedication

For my family.

What we feel most has no name but amber, archers, cinnamon, horses and birds.

-Jack Gilbert

Acknowledgements

First and foremost, I thank my advisor, Vishy Iyer, for his patience and support. He has provided an excellent example of rigorous scientific thinking. I started in his lab with no computational experience and with a shaky comprehension of molecular biology –commensurate with someone who double majored in the liberal arts. My experiences in his lab have taught me how to think more critically, and I am grateful for the opportunities he has given me.

I also thank my fellow lab members for providing helpful feedback, a supportive environment, and many, many conversations that have informed and inspired. In particular, I thank Daechan Park, whose friendship, guidance, and advice have been invaluable over the course of my graduate school career. He has graciously shared his knowledge and what was initially his own project, leading to my research on the nucleosome. Additionally, I thank Anna Battenhouse, who performed all my sequencing data alignments and data preparation, and who has been an unparalleled resource for learning computational analysis. Without her help my, by comparison, feeble efforts at data analysis would not be possible.

I also thank Chris Yellman for assistance with tetrad dissections and for advice on optimal yeast growth conditions.

I thank my committee members Edward Marcotte, Claus Wilke, Kyle Miller, and Haley Tucker. I have learned much from their advice and from the variety of their perspectives.

Finally, I thank my family. My father and my mother have shaped many of my interests and values, and are directly responsible for my curiosity about the world. My

sister Aysha is an inspiration. While we often disagree, she has provided me with a touchstone when dealing with a chaotic world in which quality can be hard to measure. Finally, my cats, Urza and Chico, have taught me how to love whole-heartedly.

The Interplay between Transcription, Histone Variants, and Chromatin Structure in Eukaryotes

Dia Nicholson Bagchi, Ph.D.

The University of Texas at Austin, 2016

Supervisor: Vishwanath R. Iyer

Transcription is a fundamental process necessary for life. In Eukaryotes this process is shaped and constrained, in part, by the 3D structure of chromatin –the assemblage of protein and DNA into which the genome is organized. Additionally, chromatin itself is reorganized as conditions change and different transcriptional programs are activated. Within this work, I present an exploration of the dynamic system created by this intricately intertwined regulation between chromatin structure and transcriptional outputs.

In Chapter 1, I begin with a review of the determinants of direction in the initiation stage of Eukaryotic transcription. The process of initiation involves numerous forms of regulation, including chromatin based. The next three chapters investigate different aspects of the nucleosome, which has been the primary topic of my research. Chapter 2 presents an overview on researching the nucleosome in the yeast *Saccharomyces cerevisiae*. Chapter 3 examines the connections between H2A.Z and transcription. Here, I challenge the generally accepted model of H2A.Z incorporation at the +1 and -1 nucleosomes hedging the transcription start site. Chapter 4 focuses specifically on perturbations to nucleosomal structure produced either from gene deletions or in response to environmental changes. Finally, I conclude by summarizing

my findings and with a general discussion of questions in the field that remain to be explored.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

The central dogma of biology as proposed by Francis Crick (Crick, 1970) outlines the canonical view of the flow of genetic information. DNA serves as a storage molecule in which information is encoded in nucleotides (composed of the bases A, T, G, and C). This information is then transferred from static but stable DNA to the less stable but mobile RNA via the process of transcription. The information contained within discreet RNA messages is then converted into polypeptide form via the process of translation. Proteins, the end product of these processes, have structural and catalytic functions and perform the various functions of life. This canonical view of the order of things in molecular biology has since been modified and amended to incorporate numerous new findings. For example, the discovery of reverse transcriptases in retroviruses has necessitated adding a path to the model through which information initially stored in RNA can be hard coded into DNA (Baltimore, 1970, Temin and Mizutani, 1970). Similarly, RNA has proven itself capable of carrying out catalytic (Lewin, 1982, Noller et al., 1992) and structural functions (Fang et al., 2015, Yan et al., 2016).

Information can also be encoded within the structure of chromatin. By controlling access of the transcriptional machinery to DNA, chromatin fine-tunes the quantity of information produced. By changing chromatin structure in ways that encourage alternate transcription start site, exon, or polyadenylation site usage (Blazie et al., 2015, Brown et al., 2012, Haberle et al., 2014), target molecules for degradation, or altogether prevent transcription, the quality of information can also be modulated. The ability of proteins to change the "message" encoded in DNA relies on the addition of covalent modifications, the inclusion of histone variants, and the ability of chromatin remodelers to mobilize nucleosomes, among many other integrated layers of chromatin

related functions. As part of the "epigenome" these changes can also be heritable from generation to generation.

In metazoans, chromatin structure varies between tissues types, as does transcriptional output, and both can be used to determine a cell type of origin (Danielsson et al., 2015, Snyder et al., 2016). These studies demonstrate that chromatin and transcription are intrinsically related, and that they unite to determine cell function. Within this dissertation I explore this inter-relatedness.

CHAPTERS

Chapter 1: The Determinants of Directionality in Transcriptional Initiation[1]

"And thus do we of wisdom and of reach,
with windlasses and with assays of bias,
by indirections find directions out."

-Polonius, *Hamlet, II, i*

INTRODUCTION

A new paradigm has emerged in recent years characterizing transcription initiation as a bidirectional process, encompassing a larger proportion of the genome than previously thought. Past concepts of coding genes thinly scattered among a vast background of transcriptionally inert noncoding DNA have been abandoned. A richer picture has taken shape, integrating transcription of coding genes, enhancer RNAs, and various other noncoding transcriptional events. This review attempts to give an overview of recent studies detailing the mechanisms of RNA Pol II-based transcriptional initiation and discuss the ways in which transcriptional direction is established, as well as its functional implications.

WHAT IS BIDIRECTIONAL TRANSCRIPTION?

The determinants of transcriptional initiation are intricate and interwoven. What is clear from the high proportion of the human genome that is transcribed (estimated at 60%) in comparison to the small proportion that is coding (2%) (Consortium et al., 2012, Djebali et al., 2012) is that transcriptional processes involve much more of the genome

[1] A version of Chapter 1 has been previously published in: 14. Bagchi, D.N. and Iyer, V.R. (2016) The Determinants of Directionality in Transcriptional Initiation. *Trends Genet*.

than was once thought. High-resolution analyses and detailed catalogs of transcription start sites (TSS) obtained using next-generation sequencing methods have shown that transcription initiation frequently occurs in both directions from a given promoter region (Preker et al., 2011, Xu et al., 2009). These studies have raised the question of whether transcription initiation is an inherently bidirectional or unidirectional process. In one model, biases in the direction of transcription arise as emergent properties from the complex regulatory restrictions placed upon inherently bidirectional promoter elements. In an alternative model, transcription at its core is unidirectional, with the appearance of bidirectionality arising due to the adjacent placement of individual unidirectional core promoters in opposite orientations. In the latter model, the similar needs of two separate gene promoters to coordinately regulate transcription factor (TF) recruitment might select for divergent transcript orientations. Transcription occurring in two directions from a single core promoter and divergent transcription originating from two distinct core promoters have not always been well distinguished in the literature. The conflation of these two categories has led to some ambiguity. Here we refer to transcription arising from a core promoter in opposite directions as bidirectional, whereas transcription of two outward facing transcripts from independent core promoters is termed divergent transcription. To some extent, the terminology that researchers in the field adopt depends on variability in the definitions and size estimates of what constitutes a promoter and how far divergent genes may lie from one another. In this review we discuss the evidence for each model to illustrate the current understanding of transcriptional initiation, and also consider the related issue of sense and antisense transcription at genes. Ultimately, we suggest a more nuanced view of promoters as non-directional, conducive regions of DNA prone to the occurrence of an open chromatin structure, the transcriptional potential of which is channeled either bidirectionally or unidirectionally in a context dependent

manner. The regulatory constraints of the various layers of regulation then work additively to produce specific transcriptional states (Figure 1.1).

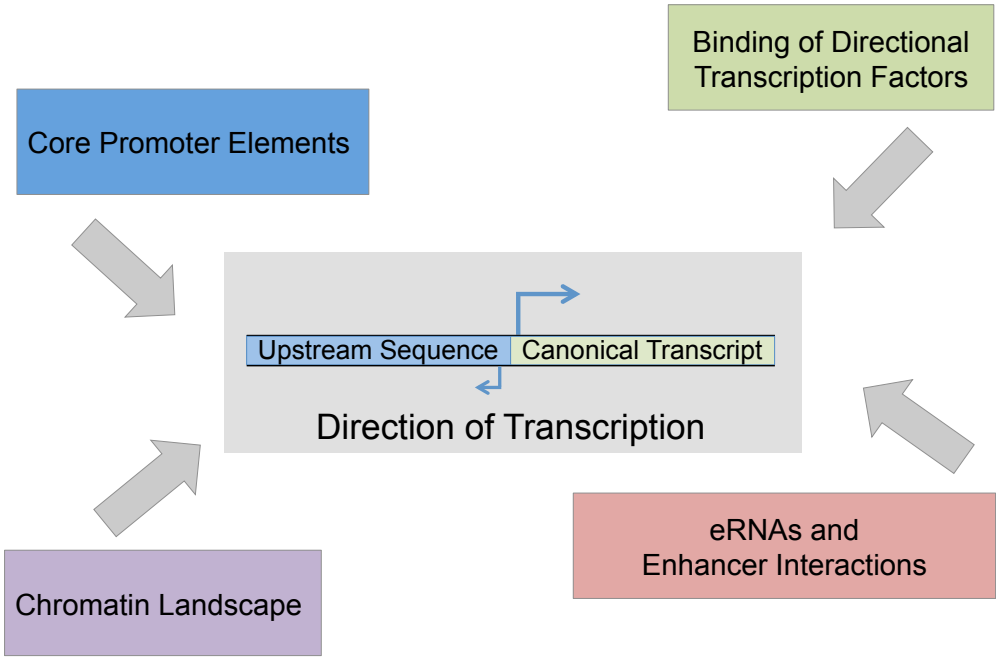

Figure 1.1: The determinants of directionality.

The schematic depicts the different factors that go into establishing directionality at a transcriptionally permissive site.

ASSAYING THE TRANSCRIPTOME

Methods

Recently, considerable effort has been directed towards using high-resolution methods to define the total RNA pool produced via transcription. High resolution methods to define the transcriptome have revealed that transcription initiates not only in the expected location downstream of promoters, but also within promoter regions

upstream of coding sequences and bidirectionally at active enhancers. Both these sources of noncoding transcription generally produce short unstable RNAs that are rapidly degraded through transcriptional termination and targeted degradation processes (Almada et al., 2013, Ntini et al., 2013). Transcription has also been observed to originate within transcript bodies (Kaplan et al., 2003), and from the 3' ends of genes in antisense orientation (Gu et al., 2015)(Figure 1.3). Numerous techniques have been used to detect nascent transcripts (Table 1). Unstable transcripts can be identified when RNA degradation pathways are inhibited, causing the persistence of unstable RNAs (Core et al., 2014, Preker et al., 2008). These experiments have been used to interrogate the genomic sites of transcriptional initiation and classify them broadly into 3 types based on their bidirectional potential: stable/stable, stable/unstable, and unstable/unstable. These categories reflect the functional directionality of a promoter but don't specify whether initiation actually occurs in both directions. For an in-depth account of the various noncoding transcripts that have been described and the techniques that have been used to identify them, see the review by Wei et al (Wei et al., 2011).

Technique	Method	Results	References
RNA Pol II ChIP Seq	Chromatin is fragmented, Pol II is immuno-precipitated, and the interacting DNA is sequenced.	Identifies DNA that is bound by RNA Pol II in a non-strand specific fashion at ~200 bp resolution.	(Barski et al., 2007, Mikkelsen et al., 2007)
cap analysis of gene expression (CAGE)/ SMORE-seq/ TIF-seq	RNA is treated with 5' cap-specific enzyme tobacco acid pyrophosphatase (TAP) and subjected to cDNA sequencing.	Identify 5' ends from stable capped RNAs at single nucleotide resolution.	(de Hoon and Hayashizaki, 2008, Park et al., 2014a, Pelechano et al., 2013)
global run-on sequencing (GRO-seq) and GRO-cap	Run on assay, which restarts RNA Pol II in vitro in the presence of a labeled nucleotide (BrUTP) in order to purify nascent RNA.	Identify nascent RNAs and post initiation pause sites at ~50 bp resolution. Since transcription is restarted in vitro, it can detect unstable transcripts, which would normally be rapidly degraded in vivo.	(Core et al., 2014, Core et al., 2008, Kruesi et al., 2013)
precision nuclear run-on and sequencing (PRO-seq)	Run on assay, which restarts RNA Pol II in vitro using biotin-labeled ribonucleoside triphosphate analogs. By supplying only 1 of the 4 nucleotides at a time, run-on transcription is limited and resolution is improved over GRO-seq.	Identifies nascent RNAs and post-initiation pause sites at <50 bp resolution. Since in GRO-seq type assays transcription is restarted in-vitro, it can detect unstable transcripts, which would normally be rapidly degraded in vivo.	(Kwak et al., 2013)
native elongating transcript sequencing (NET-seq)	RNA Pol II associated RNA is purified, and the associated RNA is sequenced.	Identifies nascent RNA at single nucleotide resolution. When combined with CTD phosphorylation specific immunoprecipitation, different populations of RNA can be identified based on the modification status of their transcribing RNA Pol II.	(Churchman and Weissman, 2011, Mayer et al., 2015, Nojima et al., 2015)
RNA-seq combined with inhibition of RNA degradation pathways	Different components of various RNA degradation pathways are inhibited (such as exosome components) to enable the isolation of unstable RNAs.	Identifies RNAs regardless of stability. Resolution is variable depending on the RNA seq method employed.	(Ntini et al., 2013)

Table 1.1: Genomic techniques for assaying transcriptional initiation

eRNAs

In mammalian cells, transcriptional activity at enhancers is widespread and dynamically regulated, generally producing unstable transcripts in both directions when actively functioning as an enhancer (Andersson et al., 2014). It should be noted though that the majority of putative enhancers identified by chromatin profiling have not been experimentally validated as being functional. A small subset of enhancers produce stable long noncoding RNAs (lncRNAs) as one of their transcript pairs. Just as the promoter-directed effects of enhancers are cell type specific and developmental timing specific, the RNAs that they produce also occur in similar waves. During changing conditions or cell states, enhancer RNA (eRNA) production is the most rapid and salient transcriptional response, preceding even the transcription of TFs in response to the change (Arner et al., 2015). The question of whether the majority of eRNAs are functional remains open. In some specific instances eRNAs have been shown to be important for the function of enhancers. Post transcriptional knockdown of a handful of eRNAs has revealed cases where they are necessary for enhancing transcription at interacting genes (Melo et al., 2013) and for promoter-enhancer loop formation (Hsieh et al., 2014). However, there are also many instances where knockdown of these eRNAs does not inhibit the function of the enhancer (Hah et al., 2013). On average, however, eRNA transcription is a good predictor of enhancer activity (Andersson et al., 2014). The potential functions of eRNAs have been discussed in a review by Li et al (Li et al., 2014).

PROMPTs

Within promoter regions, noncoding RNAs termed PROMPTs (promoter upstream transcripts, Figure 1.3) have been detected after depletion of components of the exosome, an RNA degradation complex (Preker et al., 2011). Similar transcripts have been noted by other groups and have been termed bidirectional noncoding RNAs (BNCs)

(Park et al., 2014a), cryptic unstable transcripts (CUTs) (Xu et al., 2009) or stable unannotated transcripts (SUTs) (Wei et al., 2011, Xu et al., 2009). In mammalian cells, PROMPTs have been observed to be transcribed in both the sense and antisense directions (Preker et al., 2008, Seila et al., 2008). These transcripts are generally transcribed by RNA polymerase II (RNA Pol II) but can originate upstream from Pol I and Pol III transcribed genes also (Preker et al., 2011). Antisense PROMPT transcription has been reported to be correlated (Preker et al., 2008) and anti-correlated (Preker et al., 2011) with downstream coding genes. Skewing in initiation direction may reflect trade-offs where the presence of activated open chromatin generally recruits more of the transcription machinery, but also where a transcript's abundant expression may monopolize the pool of available RNA Pol II. In contrast to stable mRNAs, most PROMPTs and eRNAs are depleted for 5' splice sites (Almada et al., 2013) and enriched for polyadenylation sites (Ntini et al., 2013), features which target them for early transcriptional termination and degradation. While the majority of PROMPTs are rapidly degraded, some stable noncoding transcripts produced from promoter regions have been shown to be functional (Albrecht and Orom, 2015, Zhou et al., 2015). Some promoter transcripts are reproducibly observed in specific tissues, cell lineages, and cancers, while others are ubiquitous (Balbin et al., 2015).

The similarity between PROMPTs and eRNAs supports the characterization of promoters as a specific type of a general class of origin of transcription, one in which a stable transcript with coding potential is produced. It has been suggested that promoters and enhancers should be viewed as a unified category of transcriptional initiation sites which are differentially regulated (Andersson et al., 2015b, Preker et al., 2008, Seila et al., 2008). Generally, eRNA transcription occurs bidirectionally, with both directions producing roughly equivalent levels of RNA (Andersson et al., 2014). In contrast, while

it is likely that most promoters produce PROMPTs, or antisense transcripts, transcription is generally skewed towards the sense direction(Balbin et al., 2015). While depletion of exosome components leads to increases for both eRNAs and PROMPTs, eRNA increases are significantly higher (Andersson et al., 2014). In some cases, intragenic enhancers can produce multi-exonic enhancer RNAs (meRNAs) which are spliced and polyadenylated just like coding genes, but are unlikely to have coding potential (Kowalczyk et al., 2012). Interestingly, not only are enhancers being recognized as resembling promoters, but promoters have been characterized with enhancer functions. Like enhancers, promoters often interact with other promoters, and in these cases, can have enhancer-like effects on their interacting partners (Kowalczyk et al., 2012, Leung et al., 2015). In this context, both elements should be regarded as sites of transcriptional initiation that are differentially characterized by the types of transcripts they produce.

CORE PROMOTER ELEMENTS WORK SYNERGISTICALLY TO ESTABLISH TRANSCRIPTIONAL DIRECTIONALITY

TATA box containing promoters

Core promoter elements are vital components in determining whether transcriptional initiation occurs and the direction in which it occurs. In bacteria the asymmetric nature of the -35 and -10 sequences recognized by the sigma factor convey directionality. In Eukaryotes, however, core promoter elements which recruit RNA Polymerases to initiate transcription come in a variety of flavors, exhibit far less conservation, and can be hard to identify. The most widely recognized of these sequences, the TATA box, has often been regarded as a directional element, in part due to a strong bias in its appearance at sites of asymmetric, directional transcription (Park et al.,

2014a). In humans, TATA box like features occur at about 29% of unidirectional promoters compared with only about 9% of bidirectional promoters (Trinklein et al., 2004). Some experiments investigating TATA box function suggest that the TATA sequence orientation matters. For example, inversion of the TATA box in the yeast *Saccharomyces cerevisiae HIS4* promoter in vivo causes a failure of *HIS4* transcription (Nagawa and Fink, 1985). However, a larger amount of data supports the notion of a bidirectionally competent TATA element, even in the context of a specifically asymmetric TATA sequence. The existence of TATA box containing promoters that show bidirectional transcription indicates that the TATA box itself is not necessarily directional in nature. Additionally, examples of inverted TATA elements can be found in natural genomic contexts (Huang et al., 1996). In vitro transcription experiments using nuclear extract from the amoeba *Acanthamoeba castellanii* have demonstrated that isolated TATA boxes support bidirectional transcription, while addition of an upstream TBP promoter element (TPE) stimulates transcription downstream of the TATA and prevents transcription upstream of the TPE (Huang et al., 1996) (Figure 1.2). In yeast, the TATA box is able to promote transcription in both orientations but the orientation does affect the level of transcriptional output (Lubliner et al., 2015). These results have also been shown to apply to mammalian transcription, as inversion of asymmetric TATA boxes in reporter plasmids transfected into human cells still produced transcriptional activation of the downstream genes they regulated (Xu et al., 1991). Since TATA elements that are strongly conserved are likely to produce at least weak TATA elements in the opposite orientation, this result is not wholly unexpected. These experiments indicate that the position of the TATA box in relation to other promoter elements is more important for determining directionality than the TATA box orientation itself.

If TATA elements have the potential to stimulate bidirectional transcription, what then accounts for the significant enrichment of TATA boxes at sites of unidirectional transcription? In yeast, TATA-containing genes tend to exhibit either high or low expression levels, are enriched for genes up-regulated in response to environmental stress, and are depleted for housekeeping genes (Basehoar et al., 2004). Additionally, TATA boxes have been associated with tissue specific promoters (Carninci et al., 2006, Engstrom et al., 2007). If TATA-regulated genes are less likely to be constitutively expressed, it is possible that divergent transcription occurs from these promoters only under specific conditions. The promoters of genes localized near telomeres are also enriched for TATA boxes (Basehoar et al., 2004). This could indicate a greater dependence on regulation by TATA elements for genes that experience heterochromatin based repression. The expression levels of these genes may need to be more dynamically regulated, as telomere-adjacent gene regions show less evolutionary conservation compared to centrally located regions (Kellis et al., 2003). A quick look at the yeast genome shows that divergent transcripts governed by TATA containing promoters tend to have an increased distance to the upstream TSS. This space could accommodate a greater number of core promoter regulatory elements. These results support a model where the regulatory requirements of TATA containing genes are more variable and demanding, selecting for their independent regulation from potential upstream transcripts.

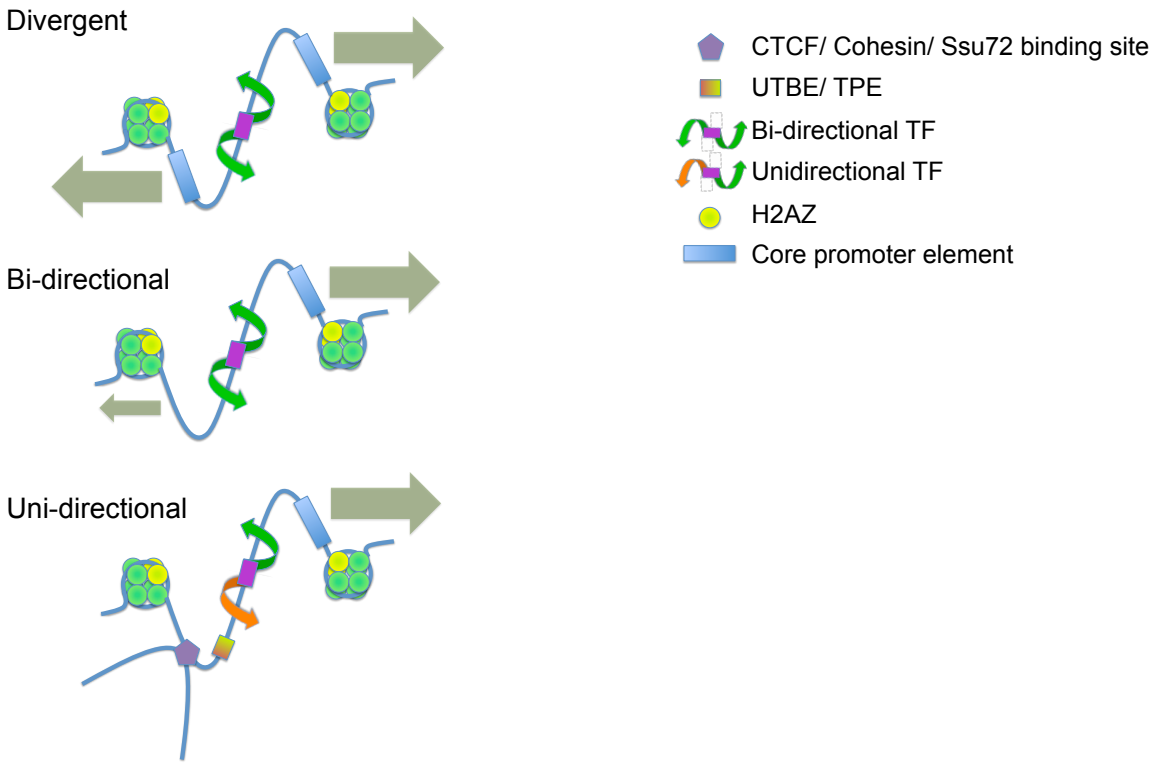

Figure 1.2: Contributors to divergent and unidirectional transcription within an NFR.

For head-to-head coding genes, most transcripts arise from two separate core promoter elements. Bidirectional transcription from a single core element may be characterized by unstable transcripts in the antisense direction in the case of coding genes, and in both directions in the case of enhancers. Finally, a subset of promoters show predominantly unidirectional transcription. As the majority of TFs have bidirectional activities, the number of truly unidirectional promoters may be relatively small. These categories may overlap and vary depending on differential conditions and tissue types. In each case, the presence or absence of transcription is subject to a variety of secondary regulation including TF expression and binding, CTCF and cohesin mediated looping, Ssu72 mediated 5' to 3' gene looping, and modifications to H2A.Z which may further promote or antagonize the progression of RNA Pol II.

Core promoter elements which establish direction

Core promoters lacking TATA boxes also impart directional preferences to the transcription originating from them. Unidirectional transcription (as measured by

luciferase reporter assays) from the insulin-degrading enzyme (IDE) gene promoter is achieved through the presence of an upstream transcription blocking element (UTBE) in both mice and humans (Zhang et al., 2013a, Zhang et al., 2013b)(Figure 1.2). In humans this element has been mapped to a specific sequence between -318 and -304 relative to the transcription start site. The downstream core promoter in isolation can promote transcription in both the sense and antisense directions. However, in the presence of a UTBE, antisense transcription is abrogated (Zhang et al., 2013b). This UTBE also has the ability to block sense transcription of SV40 transcripts when placed downstream of the SV40 promoter (Zhang et al., 2013b), highlighting the ability of this promoter element to regulate the initiation of transcription in both directions. In a more expansive study, when promoters cloned into luciferase reporters in both orientations and transfected into 4 different cell lines were assayed, some showed strong directional preferences that varied based on cell line. There was evidence that the divergent transcripts competed for the same transcriptional machinery, as deleting one TSS often increased activity from the oppositely oriented one (Trinklein et al., 2004). This indicates that the directionality associated with some promoter sequences is not solely due to intrinsic sequence factors, but relies on interactions with trans-acting cell specific factors.

Characterizing bidirectional versus unidirectional promoters

Some core promoter elements are more likely to be associated with regions of bidirectional transcription. In particular, bidirectional promoters are more likely to have a higher GC content and to fall within CpG islands. In humans, about 90% of bidirectional promoters are found within CpG islands compared to only about 45% of unidirectional promoters (Yang and Elnitski, 2008). Other elements slightly more common in unidirectional promoters include the downstream core promoter element

(DPE) and the initiator (Inr), while the CCAAT box is almost twice as likely to occur in bidirectional promoters (Trinklein et al., 2004). Characterization of the enhancer-associated unstable/unstable TSSs reveals that they display a lower CpG frequency than other bidirectional transcripts. However, they have core promoter elements (including TATA boxes and Inrs), predominantly bind the same TFs, and exhibit a canonical TSS structure with a nucleosome-depleted region (NDR) bordered by two well-positioned nucleosomes (Core et al., 2014) suggesting a similar mode of transcriptional regulation as promoters.

Despite the demonstrated bidirectional potential of many core promoter elements, only 10% of human coding genes are divergently oriented with transcription start sites (TSSs) less than 1000 bp apart (Trinklein et al., 2004). This is in contrast to yeast where approximately half of all genes are divergently oriented, and where distances between TSSs are constrained by a compact genome and high coding percentage (Chang et al., 2012). Divergent transcription is largely absent from *D. melanogaster* promoters, which show a larger number of directional motifs, but is prevalent at their enhancers (Core et al., 2012). Surprisingly, only 5% of *D. melanogaster* promoters contain a TATA box (FitzGerald et al., 2006) and these promoters generally display a strictly conserved distance between the TATA, the initiator (INR), and downstream promoter elements (DPE), which is important for promoter function and likely for directional enforcement (Kutach and Kadonaga, 2000).

For higher eukaryotes, bidirectional transcription seems to be most closely associated with the production of noncoding RNA in at least one direction. There is some debate concerning whether human promoters are inherently unidirectional or bidirectional (Andersson et al., 2015a, Duttke et al., 2015). To some extent, the answer to this question rests on which cell types are being examined, the thresholds used to call

antisense transcription and the exact methods of detection. Independent studies support the idea that bidirectional transcription is most often a result of separate core promoter elements flanking a NDR. However, there is a great deal of plasticity in what sequences are termed core promoter elements. For divergent transcripts with well defined core promoter elements, it may be that despite the potential for bidirectional activity at individual core elements, the majority of these transcripts face some selective pressure to independently regulate expression, favoring the maintenance of individual promoters which retain greater potential for independent regulation.

At sites of bidirectional transcription in yeast, two pre-initiation complexes (PICs) generally flank the NDR in an inverted orientation (Rhee and Pugh, 2012). Similarly, in mice, bidirectional transcription involves the formation of 2 distinct PICs, more TF binding, a larger, more distinctly defined NDR, and on average higher gene expression (Scruggs et al., 2015). When only one core promoter is present however, an explanation is needed for the asymmetric binding of RNA Pol II on opposite strands, which is needed to allow for bidirectional transcription (Figure 1.2). Non-consensus binding of PIC components to the NDR has been suggested to explain binding of general TFs to promoter regions without any discernable recognition sequences (Afek and Lukatsky, 2013a). Such binding was promoted by the TF Reb1 in yeast and inhibited by CTCF in human cells (Afek and Lukatsky, 2013b). NDRs with two PICs are further characterized by a greater occupancy of the -1 nucleosome (Afek and Lukatsky, 2013b), suggesting that the chromatin structure at the NDR promotes opening of a transcription bubble and is conducive to bidirectional transcription, with PICs forming in both directions at the edges of the flanking nucleosomes. Secondary regulation may then generate a predominant direction of transcription either through pre- or post- initiation regulation. The fact that promoters can be unidirectional in some tissue types and bidirectional in others (Balbin et

al., 2015) supports the notion that core promoter sequences allow for bidirectional transcription but that this capacity is then regulated by secondary mechanisms such as cell-type specific TFs that promote either one or both transcripts in the pair in response to the different needs of the cell.

TRANSCRIPTION FACTORS CAN MODULATE THE DIRECTIONALITY OF TRANSCRIPTION INITIATION SITES

Akin to core promoter elements, transcription factor binding sites mediate important interactions between the transcriptional machinery and trans-acting factors that are often necessary for transcription to occur. The binding motifs for a number of TFs, including NF-Y, Nrf-1, YY1, GABP, MYC, E2F1, and E2F4 are overrepresented in bidirectional promoters (Lin et al., 2007). These factors may act as determinants of transcriptional direction, not just passively associate with it. For example, the introduction of a GABP binding site into unidirectional promoters caused the appearance of bidirectional transcription in 67% of tested promoters (Collins et al., 2007). Further work will need to be done to determine whether insertion of motifs for other transcription factors can unilaterally change the type of transcriptional initiation arising from promoter regions, and the mechanistic underpinnings by which such effects occur. In addition to coding gene promoters, TFs are also associated with other origins of transcription. Sites producing unstable/unstable transcript pairs in B-cell derived lines show histone modifications typical of enhancers (high levels of H3K4me1) and are preferentially bound by the immune specific transcription factor PU.1 (known for binding the purine rich PU box element(Core et al., 2014). Sites producing two stable divergently arranged coding transcripts show enrichment for GABP (GA binding protein) localization (Core et al., 2014). Finally, sites producing stable transcripts only in one direction are associated

with CTCF binding, which likely involves modulating chromatin structure (Bornelov et al., 2015, Core et al., 2014).

The notion of pioneer transcription factors raises the prospect of this special group of TFs being implicated in modulating directionality. Pioneer transcription factors possess the ability to open up local chromatin conformation upon binding their motifs, thereby, facilitating the binding of additional settler TFs in a cooperative and hierarchical fashion (Soufi et al., 2012, Zaret and Carroll, 2011). The dependence on preliminary binding of pioneer TFs has been proposed to, in part, account for the very low (Joseph et al., 2010, Kaplan et al., 2011) fraction of binding sites within genomes that are actually occupied by TFs (Sherwood et al., 2014, Zaret and Carroll, 2011). Underlining their importance in determining transcriptional outcomes, pioneer TFs have been implicated in cell fate specification and cell reprogramming (Drouin, 2014, Iwafuchi-Doi and Zaret, 2014). While the majority of pioneer TFs open chromatin on both sides of their motifs, several (including Creb/ATF, Klf/Sp, NFYA, and Zfp161) open chromatin in a directional manner (Sherwood et al., 2014). These directional motifs represent a plausible mechanism by which transcription from inherently bidirectional core promoter elements is converted to unidirectional activation in genomic contexts (Figure 1.2).

THE CHROMATIN LANDSCAPE AND TRANSCRIPTIONAL DIRECTIONALITY

Histone variants

By altering the chromatin landscape in the vicinity of the TSS, transcription in either direction can be promoted or prevented. In particular, the +1 nucleosome in the direction of transcription plays an important role by presenting a barrier to the progression of RNA Pol II. This barrier can be lowered by incorporation of the histone

variant H2A.Z which has been proposed to destabilize the nucleosome and allow progression of RNA Pol II (Bonisch and Hake, 2012, Jin and Felsenfeld, 2007, Weber et al., 2014). The CCAAT box binding protein NF-Y has been found to be critical for H2A.Z deposition at the +1 nucleosome of cell cycle dependent promoters (Gatta and Mantovani, 2011), underscoring a connection between core promoter sequence and H2A.Z recruitment. Research into the covalent modifications present on H2A.Z's N-terminal tail has begun to shed light on exactly how it modulates chromatin accessibility downstream of the promoter. Mono-ubiquitination of H2A.Z has been shown to induce transcriptional repression (Sarcinella et al., 2007), and its de-ubiquitination is necessary for the activation of androgen receptor mediated genes (Draker et al., 2011). By contrast, acetylated H2A.Z is associated with increased levels of gene expression (Hu et al., 2013).

Schizosaccharomyces pombe cells lacking H2A.Z exhibit increased antisense transcription, implying a role for the histone in transcriptional repression (Zofall et al., 2009). However in *Saccharomyces cerevisiae*, H2A.Z incorporated at the 3' ends of gene bodies promoted overlapping antisense transcription (Gu et al., 2015). It is not clear if these studies reveal a difference in the role of H2A.Z in these two species, or if they reflect different covalent modification states of H2A.Z promoting different functional outcomes. But they do suggest that altering H2A.Z incorporation at the +1 and -1 nucleosomes may be a mechanism through which cells establish direction by modulating the permissibility of nucleosomes bordering the NDR to polymerase progression, thereby promoting or inhibiting transcription in each direction from the TSS. Another histone variant incorporated into NDR proximal nucleosomes, H3.3, is also recruited to promoters and enhancers during transcriptional activation and causes destabilization of nucleosome structure (Chen et al., 2013). At enhancers, asymmetric H2A.Z and H3.3 incorporation levels are associated with asymmetric Pol II enrichment (Won et al., 2015).

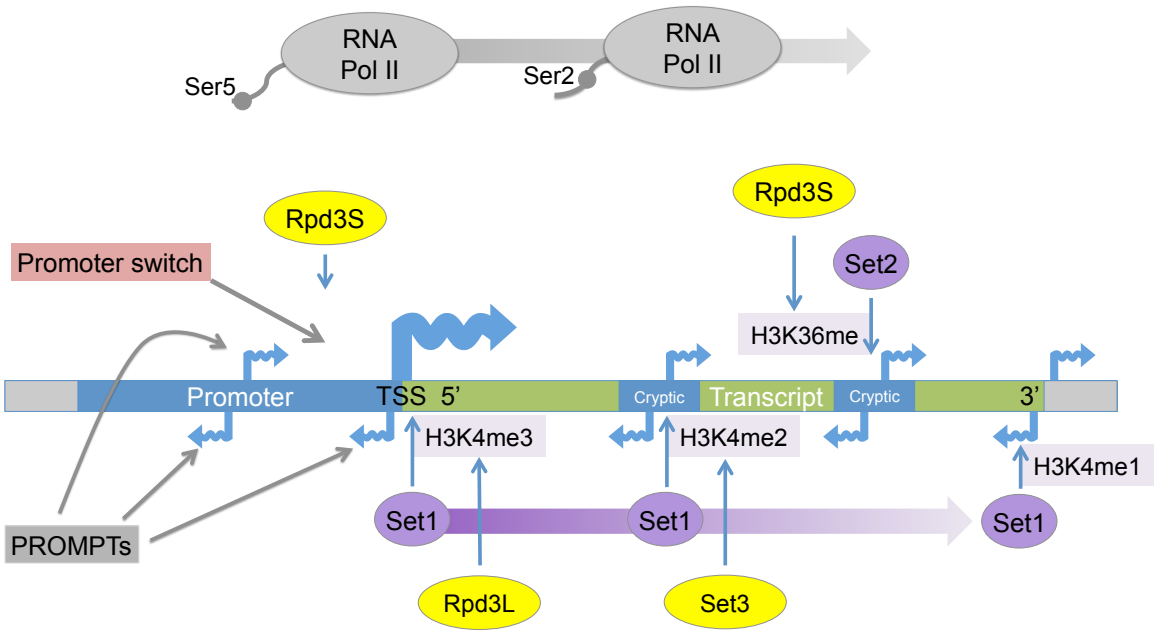

Figure 1.3: Regulation by covalent chromatin modifications and noncoding transcription.

The interplay between histone methyltransferase (purple) and histone deacetylase (yellow) activities influences noncoding transcription. When actively transcribing, RNA Pol II is phosphorylated on its CTD first at serine 5 and then at serine 2. Set 1 creates a gradient of H3K4 methylation starting with tri methylation at the 5' end and ending with monomethylation at the 3' end. Rpd3L deacetylates histones with H3K4me3 marks at the 5' ends of transcripts. Set3 deacetylates histones with H3K4me2 marks within transcript bodies. Set2 mediated H3K36 methylation takes place within transcript bodies and towards the 3' regions of the transcripts. This H3K36 modification is then targeted by RPD3S, which deacetylates histones within the 3' proximal regions of transcript bodies and prevents aberrant transcription from cryptic initiation sites within the transcript body. Rpd3S has also been implicated in promoting sense strand transcription by antisense repression upstream of the TSS.

Covalent modifications

Covalent modification of histones represents an integral component in the regulation of sense versus antisense transcription. Several chromatin remodelers and related factors have been shown to modulate antisense transcription originating from

promoters. In *S. cerevisiae*, mutations in the histone chaperone complex Chromatin assembly factor 1 (Caf1) (Marquardt et al., 2014), the histone methyltransferase Set2 (Churchman and Weissman, 2011), the histone deacetylase Set3 (Kim et al., 2012), the histone deacetylase Rpd3S (Churchman and Weissman, 2011), and the chromatin remodeler Chd1 (Hennig et al., 2012) all result in increased levels of divergent antisense transcripts. By modulating levels of antisense transcripts, these complexes help enforce directional transcription. Chd1 has also been shown to help overcome promoter-proximal stalling of RNA Pol II, facilitating productive elongation of sense transcripts (Skene et al., 2014). In the opposing direction, deletion of Hda2 or the histone methyltransferase Set1 increases sense transcription by decreasing the presence of transcribed antisense RNA (Camblong et al., 2009, Camblong et al., 2007).

Specific histone modifications characterize divergent human promoters, which are enriched for marks associated with transcriptional elongation (such as H3K4me2-3 and H3K27ac) in both the downstream and upstream directions, while unidirectional promoters lack this enrichment in the upstream direction (Duttke et al., 2015). Processes leading to asymmetric chromatin enrichment patterns across transcriptional initiation sites can be inhibitory or conducive to transcription in either direction. During transcription RNA Pol II is phosphorylated on its C terminal domain (CTD) first at serine 5 (Ser5P) near the 5' end of genes and then at serine 2 (Ser2P), allowing RNA Pol II to enter into the elongation phase of transcription, and establishing correct chromatin structure across transcribed regions, as many chromatin remodelers interact with the CTD (Srivastava and Ahn, 2015). These chromatin remodelers prevent aberrant transcription from taking place within gene bodies after the perturbations to chromatin structure that occur concomitant to transcription.

The Set and Rpd3 histone methyltransferases and deacetylases exemplify the mechanisms by which the spread of co-transcriptional modifications can in turn regulate transcription (Figure 1.3). Histone methyltransferase Set 1 (also known as the COMPASS complex) interacts with RNA Pol II Serine5P but not RNA Pol II Serine2P and creates a gradient of H3K4 methylation across transcript bodies, starting with tri-methylation at the TSS, and ending with mono-methylation near the 3' end. These methyl marks are then targeted by two separate histone deacetylases. Rpd3L (the larger complex) recognizes H3K4me3 marks and is likely targeted to promoters to regulate initiation in a sequence dependent manner (Terzi et al., 2011). Set3 is a histone deacetylase that despite having a Set domain has not been shown to have methylase activity. It targets and requires H3K4me2 to deacetylate histones, and has been shown to repress 5' proximal cryptic transcripts (Kim and Buratowski, 2009)(Figure 1.3).

The histone methyltransferase Set2 contributes H3K36me marks to 3' transcribed regions. This mark is recognized by the histone deacetylase Rpd3S (the smaller complex), which works at primarily intragenic regions in a more sequence independent manner than Rpd3L (Carrozza et al., 2005). Rpd3S removes the acetylation that co-occurs with transcriptional elongation, thereby suppressing cryptic transcripts originating within gene bodies near the 3' end. These transcripts are thought to arise because the chromatin remodeling concomitant with transcription is perturbed, leading to increased accessibility of intragenic regions to TFs and Pol II (Lickwar et al., 2009). Rpd3S has also been implicated in the repression of antisense transcripts, thereby promoting sense strand directed transcription (Churchman and Weissman, 2011)(Figure 1.3). These chromatin remodelers demonstrate the general principle that by spreading chromatin modifications across a genomic region, transcription can be inhibited or promoted. Processes leading to asymmetric chromatin enrichment patterns across transcriptional

initiation sites could likewise be inhibitory or conducive to transcription in either direction.

Chromatin loops

A further level of chromatin-based regulation of transcriptional directionality involves chromatin loops. In general, these loops serve to increase the levels of unidirectional sense strand transcripts. Polyadenylation complex factor Ssu72 facilitates the formation of gene loops between the 3' and 5' end of genes, and loss of these gene loops leads to increased levels of divergent transcription at yeast promoters (Tan-Wong et al., 2012). In mammalian cells, CTCF and cohesin facilitate chromatin loop formation (Hou et al., 2008, Tark-Dame et al., 2014), and display a biased association with unidirectional transcripts (Bornelov et al., 2015)(Figure 1.2). In these genes, CTCF and the cohesin component Rad2 are often found a short distance (60-80 bp) upstream of the TSS and their enrichment level is anti-correlated with antisense transcription. These data emphasize the importance of loop formation in directing transcription. The many epigenetic factors involved in repressing antisense transcription support a view of promoters where an inherent predisposition towards bidirectional transcription must be actively controlled.

Chromosomal looping also occurs between enhancers and promoters, and this interaction is vital for the transcription promoting activities of enhancers. Surprisingly, enhancers and promoters not only share the ability to serve as sites of transcription initiation, but also share the ability to promote transcription at locations that they physically interact with. Just as enhancers increase the likelihood of transcription at interacting promoters, TSS that physically interact with enhancers through loops stimulate the production of eRNAs (Sanyal et al., 2012). These analyses have also

revealed that in some cases promoters can function as enhancers for other promoters (Kowalczyk et al., 2012, Leung et al., 2015). In general, enhancers form fewer connections than promoters. On average, a promoter is associated with 4.9 enhancers while an enhancer is associated with 2.4 promoters (Andersson et al., 2014). The determinants of promoter/enhancer interaction specificity have been explored in a review by van Arensbergen et al (van Arensbergen et al., 2014).

FUNCTIONS OF NCRNAS

Although there are many known examples of transcriptional regulation by ncRNA or antisense RNA, it has been difficult to ascribe functional relevance to the majority of PROMPTs and eRNAs. For example, deletion of exosome component RRP6 increases upstream antisense transcription and represses sense transcription of a certain subset of genes. Whether this is through titration of TFs and RNA Pol II away from the downstream gene, or effects of the transcribed PROMPT RNAs themselves remains to be determined (Castelnuovo et al., 2014). The effects of overlapping transcription on coding gene transcription are likely to be highly context specific and depend on the direction of the overlapping transcription, the length of the coding gene, and the types of co-translational chromatin modifications. Each of these features is likely to affect the types of methylation gradients and chromatin modifications spread. As an example, for many Set3 regulated genes, upstream originating overlapping transcription places Set1 dependent H3K4me2 over promoters and causes deacetylation by the Set3 complex, thereby repressing coding gene transcription. Loss of transcription from these overlapping transcripts or from internal antisense cryptic transcripts de-represses the coding genes (Kim et al., 2012) (Figure 1.3).

Other sources of overlapping transcription also modulate the likelihood of a gene being transcribed. In particular, a group of stochastically controlled "promoter switches" regulates the genes for the class I major histocompatibility complex (MHC) receptors in mouse and humans (Anderson, 2014). Within these promoters, antisense transcription upstream of the primary transcript is generally associated with a transcriptionally off state while sense-directed transcription originating upstream of the promoter represents an on state (Figure 1.3). Changes to the direction of transcription within the switch can thereby switch the activation state of the downstream gene. Similarly, antisense transcription originating from within gene bodies and converging on the promoter is a feature found within a group of low expression genes(Mayer et al., 2015).

Many noncoding transcripts are up-regulated following growth related changes in media nutrient composition or other environmental conditions. These changes are often condition specific and consistent. These transcriptional changes are often condition specific and consistent. These observations suggest two possibilities. First, the mechanisms controlling repression of cryptic transcripts may be complex and require multiple factors to be achieved. Perhaps, any dramatic change in overall gene expression patterns has the potential to disturb their tight regulation. The factors that must be employed to achieve this repression may differ under different conditions. The up-regulation may be transient before transcriptional homeostasis is achieved. Alternatively, the increase in cryptic transcripts could be actively regulated and functionally relevant. This would indicate that they are somehow involved in acclimation to environmental change and to the activation of appropriate transcriptional programs. It is plausible that the cryptic transcripts themselves could be functional, as tissue specific enhancers have been characterized within exons (Birnbaum et al., 2012, Ritter et al., 2012).

Some TFs have been shown to bind RNA as well as DNA (Cassiday and Maher, 2002, Sigova et al., 2015). YY1 binds RNA at promoters and enhancers. When RNA transcription occurs, the nascent RNAs can bind YY1 and increase its occupancy at the NDR, creating positive transcriptional feedback. This RNA-mediated effect has led to the hypothesis that trapping of TFs (especially ones with RNA binding domains) at the TSS by RNA may be a general mechanism by which RNAs regulate transcriptional processes. This also provides a plausible function for upstream antisense transcription in recruiting TFs to the NDR so that they may promote downstream coding gene transcription (Sigova et al., 2015). In a different form of RNA based regulation, YY1 has also been shown to interact with a long intergenic noncoding RNA (lincRNA) transcribed from its own promoter region. When Linc-YY1 binds to YY1 it can de-repress YY1 target genes by causing the eviction of YY1 and PRC2 from gene promoters (Zhou et al., 2015). Two other well characterized lncRNAs that have been shown to bind PRC2 include COLDAIR in Arabidopsis, which mediates repression of the flowering control gene FLC (Heo and Sung, 2011, Ietswaart et al., 2012), and HOTAIR in Drosophila, which regulates the HOXD locus (Rinn et al., 2007). The individual transcriptional regulatory activities of a number of different lncRNAs have been previously discussed in a review (Albrecht and Orom, 2015).

CONCLUDING REMARKS

The model that current knowledge in the field presents is one where a bidirectional core promoter element recruits the transcription machinery to an accessible NDR. Additional promoter elements, TF binding sites and chromatin features then specify the directionality of transcription originating from this location. Post-initiation

factors can then modulate the elongation and stability of transcripts. These regulatory features may vary in response to changing environmental conditions and in metazoans, according to cell type. Finally, overlapping sense transcription and 3' end originating antisense transcription can spread chromatin signatures, which further encourage or repress transcription. The transcriptional output is likely to rely on a hierarchical ordering of these factors, the complex interactions of which are organism and context dependent (Figure 1.1).

Several issues remain to be addressed. Future experiments must determine the factors that promote concordant regulation of bidirectional transcripts versus anti-correlated regulation. In particular, the two classes should be analyzed for differential TF binding motifs to assess whether correlated expression results from promoters with general TF binding sites (where more TF sites represent higher transcription in both directions), and anti-correlated expression from promoters with directional TF binding sites. There is also a need to determine whether PROMPTs, especially sense strand PROMPTs (Figure 1.3), and promoter switches arise from core promoter element-like sequences. Further, the characterization of bidirectional promoters needs to distinguish between promoters from which transcription is initiated in both directions (such as a PROMPT-coding gene pair) and from which stable transcripts are produced in both directions (such as coding gene pairs) as these two categories are often conflated. While the compact genome size and high density of coding regions in *S. cerevisiae* may necessitate bidirectional transcription, the same logic fails to explanation such orientations in the human genome where only ~2% of the genome is coding (Ng et al., 2009). It has been suggested that the bi-directional orientation of genes in humans is a result of the human genome having evolved from a more compact ancestral genome (Takai and Jones, 2004). A number of cancer associated gene pairs are transcribed

divergently in a coordinate manner, suggesting that shared regulatory requirements, and transcriptional interdependencies are what keep pairs together (Albrecht and Orom, 2015).

More work needs to be done to define how the directional specificities of specific core promoter elements and transcription factor binding sites are integrated to determine transcriptional direction. It remains to be determined which individual core promoter sequences are capable of initiating bidirectional transcription and whether they actually do so in their genomic contexts. For the majority of TFs, changing the orientation of their binding sites does not significantly change their influence on gene expression (Sharon et al., 2012), and co-varying expression levels provides further evidence that TF binding increases the likelihood of transcription in both directions. Differential regulation of outward-facing transcripts in yeast is associated with the presence of insulator-like DNA binding factors Tbf1 and Mcm1 (Yan et al., 2015). Two possibilities present themselves as a way of explaining the phenomenon of bidirectional transcription. In the first, antisense transcription is a by-product of forward transcription and an NDR, meaning that the cell may need to utilize mechanisms to prevent the accumulation of detrimental antisense transcripts. In the second, there is a function for antisense transcripts with a number of possible effects including the regulation of sense transcripts. While these functions do not need to be mutually exclusive, they could create a dichotomy between two different types of TSSs. If antisense transcription is a necessary outcome of forward transcription, a single core promoter might be the most commonly encountered situation. However, two independent promoters would allow for more precise regulation of the different transcripts. Ultimately, the regulation of transcriptional direction is accomplished by both pre and post initiation factors and integrates many

factors such as promoting and repressing the initiation of transcription, regulating elongation and termination, and targeting unstable transcripts for rapid degradation.

Chapter 2: An Introduction to Studying the Nucleosome in the Yeast *Saccharomyces cerevisiae*

"As you set out for Ithaka
hope the voyage is a long one"

-C. P. Cavafy, *Ithaka* (translated by Edmund Keeley)

CHROMATIN

The nucleosome as the basic unit of chromatin

In Eukaryotes, the need to organize a large amount of DNA[2] within the confines of a nucleus and to regulate the accessibility of this DNA, has been answered with chromatin –the complex of DNA and proteins into which the genome is packaged. Chromatin can allow DNA to be compacted by as much as 10,000 times (Jiang and Pugh, 2009). At the most basic level of organization within chromatin, the DNA composing a chromosome is packaged into nucleosomes, a structure which is often descriptively compared to "beads on a string"[3] (Olins and Olins, 1974). An individual nucleosome consists of 147bp of DNA wrapped around a core of 8 histones (2 each of H2A, H2B, H3, and H4) approximately 1.65 times (Luger et al., 1997). This association is mediated through interactions between the negatively charged DNA backbone and the arginine and lysine rich histone proteins, which bear a positive charge. An additional H1 linker histone is found outside the nucleosome core but can stabilize the nucleosome by binding adjacent to it (Allan et al., 1980). The region of DNA between individual nucleosomes is called the linker region, and is, on average, approximately 30 bp long (Shivaswamy et al.,

[2] In humans, approximately 2 meters of DNA is squeezed into the cell nucleus. 21. Bloom, K. and Joglekar, A. (2010) Towards building a chromosome segregation machine. *Nature* 463, 446-456.

[3] In the original paper they were described as "particles on a string."

2008). However, the precise range of these linkers varies between species. Areas of the genome where chromatin is tightly compacted and where, consequently, gene expression is repressed, are termed heterochromatin. In contrast euchromatin describes a chromatin structure that allows for access by components of the transcriptional machinery. In regions with active transcriptional initiation a nucleosome depleted region[4] (NDR) is a common feature, allowing for greater accessibility of DNA to transcriptional machinery. NDRs are approximately 160-170bps wide (Yuan et al., 2005) and are enriched for polyA and polyT sequences which can produce a bent structure refractory to nucleosomes (Nelson et al., 1987).

The positioning of nucleosomes within chromatin depends on numerous factors. These can include intrinsic DNA sequence binding preferences of histones, the effects of processes that necessitate the repositioning of nucleosomes, and the actions of ATP dependent chromatin remodelers, which can incorporate, evict, or reposition nucleosomes. The most well defined nucleosome arrangement is that which is found at the transcription start site (TSS). Here, a wide NDR is bordered by two well-positioned nucleosomes, which set the pattern for a periodicity in nucleosome positioning extending upstream and downstream (Yuan et al., 2005). As distance from the NDR increases the periodicity becomes less well maintained.

Histone variants

The HGNC (HUGO Gene Nomenclature Committee) database lists approximately 88 human histone genes that are not annotated as pseudogenes (Gray et al., 2016). These genes code for 52 individual histone proteins. By contrast, the 11 histone genes in the yeast *Saccharomyces cerevisiae* code for 5 canonical histones and 2 variants (H2A.Z and

[4] This region is sometimes referred to as the nucleosome-free region (NFR) as well.

CenH3). These variants may replace the canonical histones within nucleosomes (Table 2.1). The limited set of histones found within yeast means that studying histone variants in yeast is much more tractable than in humans, and that the information gained is likely to be relevant in a global fashion. By comparison, histone variants in humans are often associated with specific tissue types, or specific developmental stages, which drastically increases the complexity of studying the dynamics of histone variant incorporation.

H4 shows the least amount of amino acid sequence variation across the histone families. It is worth noting, that a number of reviews claim that there is one H4 variant in humans (Kamakaka and Biggins, 2005, Marzluff et al., 2002, Maze et al., 2014). However, the HUGO database lists two distinct species, one of which is encoded for by 14 genes, and one, which is encoded specifically by the gene HIST1H4G. A BLAST database search for the HIST1H4G amino acid sequence reveals two high sequence identity hits to predicted amino acid sequences from the common chimp and from gorilla that are supported by mRNA evidence. An alignment (using Clustal Omega 1.2.1) between the two histone variants reveals that the G type variant tends to vary more in sequence between species, but that there are a number of differences from the canonical H4 sequence that are conserved (Figure 2.1). The dbSNP database also lists an SNP occurring nearing the N terminus of the human HIST1H4G gene (dbSNP build 146 rs41266821 with an allele frequency of ~8%) that converts Valine to Alanine. A BLAST search for the consensus nucleotide sequence between human, chimp, and gorilla for the G type variant also turns up hits for a number of uncharacterized cDNAs from other species. It is unclear why this variant has been ignored by the scientific literature, but the evidence for its expression in a number of organisms and for variant alleles within humans argues for the utility of characterizing this variant in future studies.

Histone Family	# Yeast Histone Genes	Yeast Histone Protein Variants	# Human Histone Genes	# Human Histone Protein Variants
H2A	3	H2A, H2A.Z	26	19
H2B	2	H2B	20	16
H3	2	H3, CenH3	17	5
H4	2	H4	15	2
H1	1	H1	10	10

Table 2.1: Histone variants in *Saccharomyces cerevisiae* and Humans

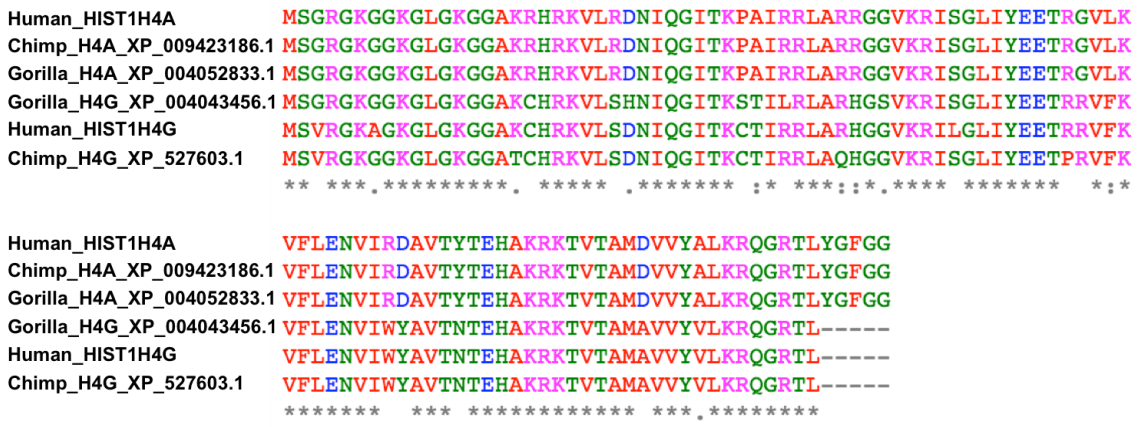

Figure 2.1: Alignment between HIST1H4A and HIST1H4G

Clustal Omega 1.2.1 based alignment between the amino acid sequences for the Human H4 genes and the common chimp (*Pan troglodytes*) and gorilla (*Gorilla gorilla*) H4 genes. The sequences for chimp and gorilla G type H4 are predicted while the A type sequences are annotated.

The regulation of histone variant H2A.Z

The histone variant H2A.Z is highly conserved, and found throughout Eukaryotes displaying sequence conservations of ~70-90% (Iouzalen et al., 1996). Performing a multiple sequence alignment (using Clustal Omega) between the *S. cerevisiae* H2A.Z protein sequence and the closest human H2A.Z ortholog, H2AFZ, reveals that they share a 68.5% identity (Figure 2.2). In contrast, the *S. cerevisiae* protein sequences for H2A.Z and the canonical histone H2A reveals that they share a 61% identity, with H2A.Z containing additional amino acids at both the N-terminal and C-terminal ends (Figure 2.3).

In *S. cerevisiae* the gene encoding H2A.Z is called Htz1. From here on in, when referring to the gene locus, I will use the term Htz1 and when referring to the histone

protein I will use H2A.Z. H2A.Z replaces the canonical H2A histone within nucleosomes at specific sites within the yeast genome, including at nucleosomes bordering the transcription start site. Replacement is catalyzed by the ATP dependent chromatin remodeler Swr1 (Mizuguchi et al., 2004). The exchange takes place by removal of an H2A/H2B dimer and replacement with an H2A.Z/H2B dimer. In the opposite direction the ATP dependent chromatin remodeler Ino80 exchanges H2A.Z/H2B dimers for H2A/H2B dimers (Papamichos-Chronakis et al., 2011)(Figure 2.4). These dynamics occur on a genome wide scale at sites of transcriptional initiation. Additionally, the incorporation of H2A.Z subunits within nucleosomes is vital for maintaining the boundary between heterochromatin and euchromatin, which occurs within the chromosome at centromeres and telomeres. Loss of H2A.Z allows heterochromatin to spread past it proper boundaries and into what should be euchromatic regions (Meneghini et al., 2003).

In most metazoans H2A.Z is essential. However, *S. cerevisiae* tolerates loss of H2A.Z with little fitness defect under optimal growth conditions (Santisteban et al., 2000). This result is somewhat puzzling given strong H2A.Z localization to sites of transcriptional initiation. However, this tolerance also gives us the unique ability to study the effects of H2A.Z loss.

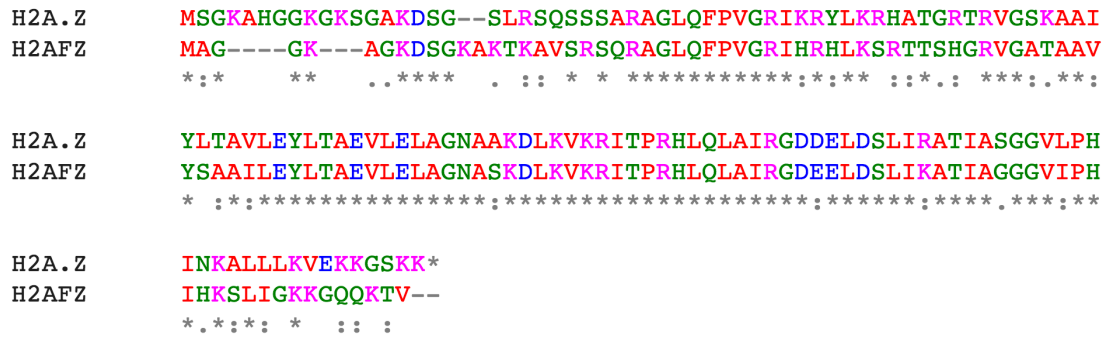

Figure 2.2: Alignment between yeast H2A.Z and human H2AFZ protein sequences

The protein sequence for H2AFZ was obtained from the Uniprot database, while the protein sequence for H2A.Z was obtain from the Saccharomyces Genome Database (SGD). Sequences were aligned with Clustal Omega.

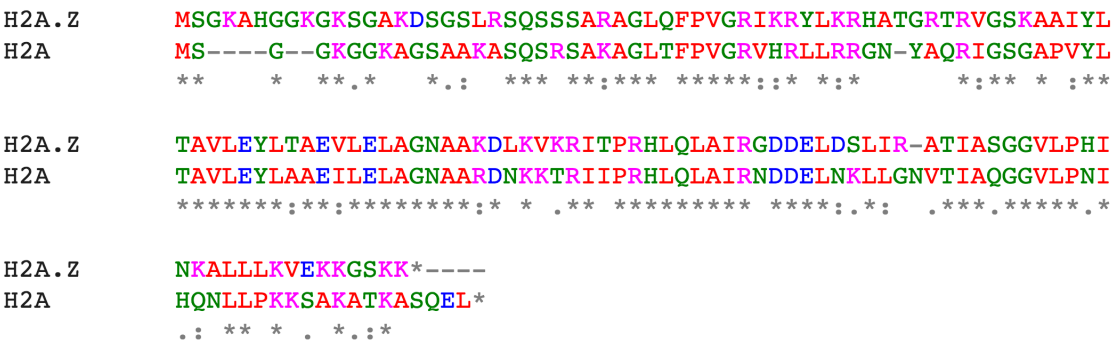

Figure 2.3: Alignment between yeast H2A.Z and H2A protein sequences

The protein sequences for H2A.Z and H2A were obtained from (SGD) and were aligned with Clustal Omega.

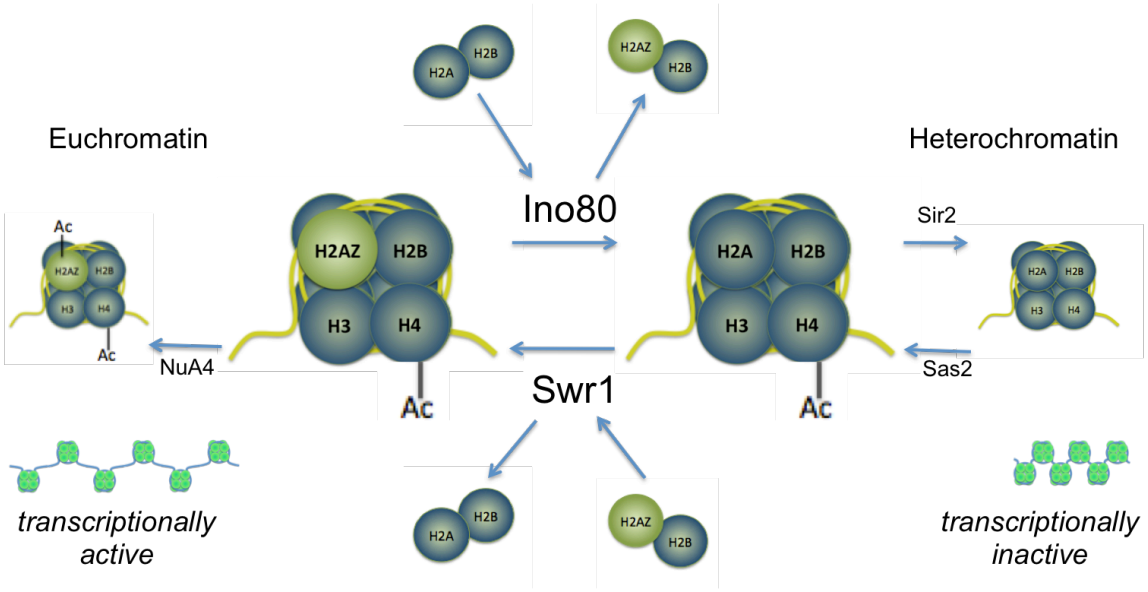

Figure 2.4: The nucleosome incorporation dynamics of H2A.Z

The chromatin remodelers Swr1 and Ino80 work in opposition to regulate the incorporation of H2A.Z into nucleosomes.

Covalent modifications

Covalent modifications are often added to the tails of histones extruding from the nucleosome. These modifications are added primarily to N-terminal tails (although some C-terminal modifications are possible). Modifications that have been described are numerous and include acetylation, methylation, ubiquitination, and phosphorylation, amongst others (Audia and Campbell, 2016). The combination of these modifications has been proposed to comprise a "histone code", functioning like a language that can be read by interpreter proteins that then perform actions based on these instructions (Strahl and Allis, 2000). Varying these modifications can determine chromatin accessibility and can recruit transcription factors to modulate transcription outputs.

Covalent modification of H2A.Z has also been described, and is primarily accomplished by the histone acetyltransferase NuA4 (Keogh et al., 2006). H2A.Z can be acetylated at Lys3, Lys8, Lys10, and Lys14. Lys14 is the most abundant of these modifications, and is associated with active gene promoters. In humans, H2A.Z can also be ubiquitinylated at Lys120, Lys121, and Lys125 by repressive complex PRC1 (Subramanian et al., 2015), and its monoubiquitylation is associated with facultative heterochromatin (Sarcinella et al., 2007).

SACCHAROMYCES CEREVISIAE AS A MODEL ORGANISM

Genome

S. cerevisiae is an extraordinarily useful model organism within molecular biology. As a Eukaryote it shares many of the same transcript and chromatin regulatory features as metazoans, but in a more simplified format. The genome is composed of approximately 12 million bp of DNA divided among 16 chromosomes and contains approximately 6000 genes. Their compact genome means that the cost of sequencing is low. The modest gene content of their genome also means that there are fewer gene homologs that must be considered when investigating the actions of particular protein families.

Unlike in humans where the majority of the genome lacks coding potential, the *S. cerevisiae* genome is densely packed with coding genes, often to the point of being interleaved. This density leaves little room for sizeable intergenic regions, and means that particular attention must be paid to transcript orientation and to the distance from TSSs to upstream transcripts when studying the structure of the NDR.

Growth conditions and strain availability

S. cerevisiae is typically grown at 30°C either in liquid yeast peptone dextrose (YPD) culture media, or on petri dishes containing YPD and agar. Doubling times are on the order of 1.5 hours. This relatively rapid growth rate enables the production of large amounts of experimental material over a relatively short time scale. The existence of numerous commercially available strain collections (including deletion collections, tagged protein collections, inducible gene expression collections, etc.) has allowed for systematic analyses and lowered the entry barrier to studying specific protein and gene functions.

METHODS

ChIP-seq

Chromatin Immuno-precipitation followed by DNA sequencing is a technique used to map protein binding sites along DNA on a genome wide basis. To achieve this, protein and DNA interactions are fixed by adding formaldehyde to live cells. DNA protein complexes are then isolated. This is done either by using an antibody that recognizes the protein of interest, or by attaching a peptide tag to the end of a protein and then pulling down the tagged protein with antibody bound beads. The DNA associated with these complexes is then purified and sequenced. Aligning these sequences to a reference genome allows determine of genomic regions for which the protein of interest has a high binding affinity under the specific laboratory conditions used.

MNase-seq

MNase-seq is a technique used to isolate mono-nucleosomal DNA from cells and, thereby, determine the consensus genome wide localization of nucleosomes within a

population. First, zymolyase is used to permeabilize the yeast cell wall. Then, micrococcal nuclease (MNase) treatment at a range of concentrations is used to degrade non-nucleosomal DNA. The structure of the nucleosome protects DNA wrapped around it from degradation by MNase. The resulting purified DNA can then be run on a gel to visualize the nucleosomal ladder produced, and to isolate the DNA gel band corresponding to mono-nucleosomal DNA. This DNA can then be sequenced and aligned to the yeast genome in order to determine sites of nucleosome localization.

MNase ChIP-Seq

By combining chromatin immunoprecipitation with micrococcalnuclease treatment, MNase ChIP-seq provides a way of pulling down nucleosomal DNA complexes that either contain specific histone variants or modifications, or that interact with specific protein factors (Wal and Pugh, 2012) (Figure 2.5). Input samples are, also, usually produced and provide data on background nucleosome occupancy levels. These control samples provide data that is equivalent to MNase Seq. By subtracting out input derived background nucleosome levels from MNase ChIP data, high resolution binding and incorporation patterns at individual nucleosomes can be produced.

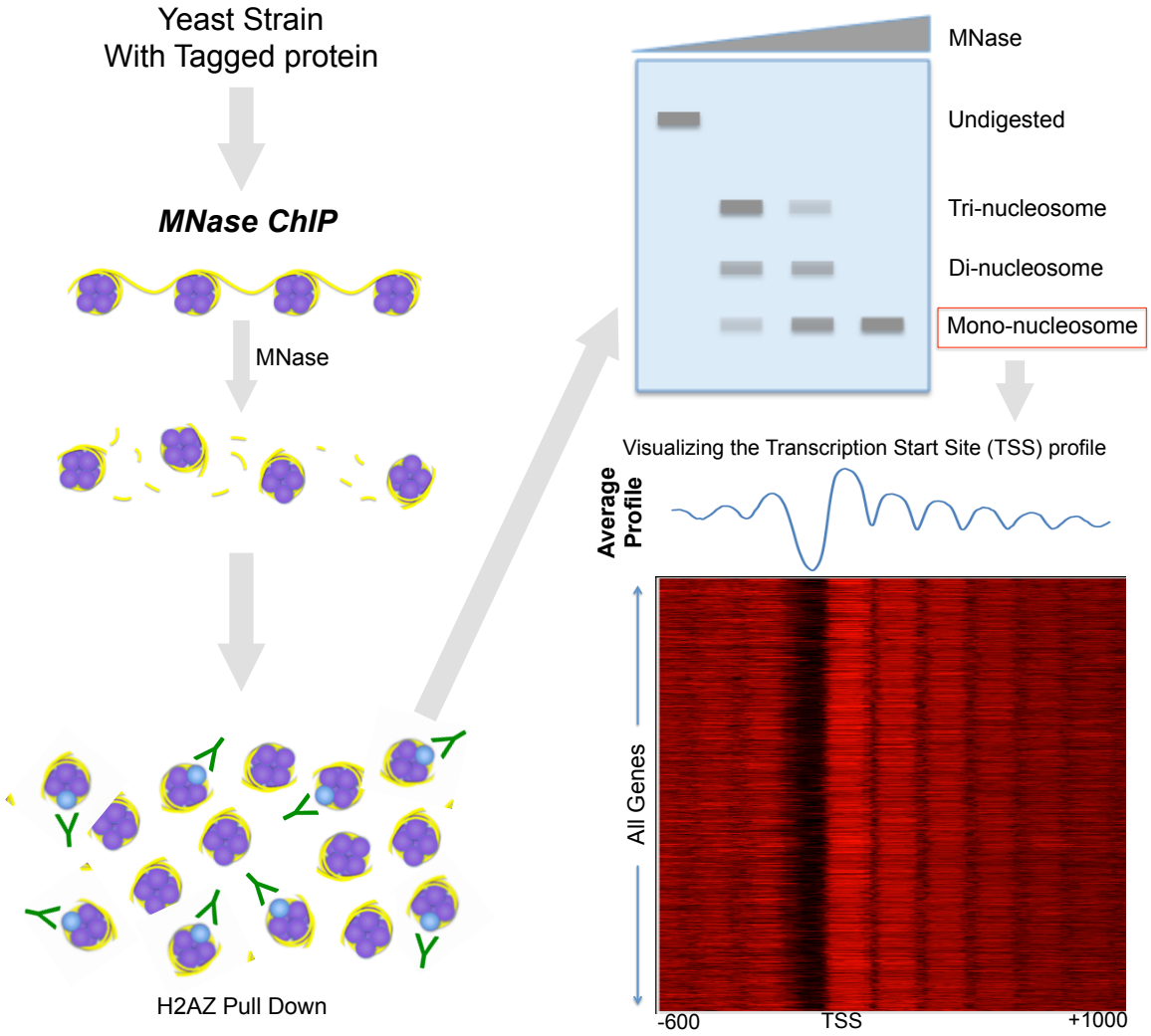

Figure 2.5: Overview of MNase ChIP-seq method

RNA-seq

RNA sequencing provides a way to assay changes in the transcriptome. Cells are grown up to an appropriate OD, and then spun down to isolate a cell pellet. This pellet is then flash frozen using liquid nitrogen in order to preserve the RNAs from degradation by RNases. When the pellet is re-suspended it is done in a buffer that inhibits RNA

degradation. In yeast, total RNA can be isolated using a hot phenol extraction method. In order to sequence mRNAs, the transcripts must first be purified away from the very high levels of ribosomal RNAs (rRNAs). While, within log phase cells, total RNA is abundant (the ratio of RNA to DNA is 50:1), rRNA accounts for 80% of the RNA, while tRNA is 15% and mRNA is 5% (Warner, 1999). Two methods can be used to increase the representation of mRNAs: 1) ribosomal rRNA depletion and 2) poly-A selection. RNAs are then fragmented and reverse transcribed into cDNAs. These cDNAs can then be used to make strand specific libraries, which are sequenced by standard DNA sequencing methods. Sequencing data can then be aligned to a genome, quantified for transcript abundance, and used in differential gene expression analysis with publicly available programs.

SMORE-seq

Simultaneous mapping of RNA ends (SMORE-seq) provides a method for determining the 5' ends of transcripts (Park et al., 2014a). This technique allows for precise mapping of transcription start sites, and, can also allow the detection of short antisense RNAs that are difficult to detect by standard RNA-seq methods. In particular, this technique can be used to identify antisense transcripts arising from between tandemly arranged genes, which have been termed BNCs (for sites of bidirectional non-coding transcription).

Chapter 3: H2A.Z as a Marker of Transcription

"We cast a shadow on something wherever we stand, and it is no good moving from place to place to save things; because the shadow always follows. Choose a place where you won't do harm - yes, choose a place where you won't do very much harm, and stand in it for all you are worth, facing the sunshine."

-George Emerson, *A Room with a View*

INTRODUCTION

The histone variant H2A.Z is generally incorporated into nucleosomes bordering the nucleosome depleted region (NDR) at the transcription start site (TSS). These two nucleosomes are referred to as +1 for the nucleosome immediately downstream of the NDR in the direction of transcription, and -1 for the upstream nucleosome. In yeast, the TSS is predominantly found at about one helical turn into the +1 nucleosome (Albert et al., 2007). Overall, the presence of nucleosomes inhibits transcription (Wasylyk and Chambon, 1979), and their presence bordering the NDR provides a plausible means of gating transcriptional output. It has been proposed, that, when H2A.Z is incorporated, it may destabilize the nucleosome structure, making its eviction more feasible, and lowering the threshold for transcriptional initiation. However, conflicting studies have provided evidence for both more and less stable structures (Zlatanova and Thakar, 2008) as well as no significant effect on stability (Thakar et al., 2009). One interesting thing to note, is that the region of H2A.Z that contacts the nucleosome core often displays less conservation, perhaps suggesting that H2A.Z's affect on nucleosome stability may be somewhat species specific (Bonisch and Hake, 2012). Ultimately, the stability of H2A.Z containing nucleosomes is likely subject, at least in part, to the covalent modifications it is endowed with, the other histones with which it partners, and the heterotypic or homotypic nucleosomes it produces.

Despite issues of stability, what seems clear is that incorporation of H2A.Z has the potential to decrease the barrier posed by the +1 nucleosome (Weber et al., 2014). Additionally, the presence of H2A.Z is associated with high nucleosome turnover rates (Dion et al., 2007). It has been argued that RNA Pol II does not actively evict nucleosomes as it transcribes, but, that it relies on fluctuations in the nucleosome core to access DNA (Hodges et al., 2009). It is possible that H2A.Z could alter the type or frequency of these fluctuations. These studies suggest a relationship with transcription that may be highly context specific, but that is nevertheless significant. Within this chapter, I explore the connections between transcription and H2A.Z's incorporation at the NDR as they are, specifically, laid out in yeast.

MATERIALS AND METHODS

Strains utilized and growth conditions

The yeast strains used for the following experiments were all from the WT haploid BY4741 background (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) and the yeast haploid deletion strain collection (Open Biosystems, now GE Dharmacon) (Winzeler et al., 1999). The primary strain used for MNase ChIP-seq experiments contained a TAP-tagged copy of the histone protein H2A.Z from the TAP-tagged protein collection (Ghaemmaghami et al., 2003). Additional Myc-Tag strains were created via transformation of 13Myc-His3MX6 cassettes amplified with primers targeted to the C-terminus coding region of Swr1 or Ino80 from the pFA6a-13Myc-His3MX6 plasmid (Longtine et al., 1998).

Yeast cells were grown up in liquid culture in yeast extract peptone dextrose (YPD) at 30°C until the cells reached a concentration measured via A600 OD of

approximately .8. The cells were then spun down in an ultracentrifuge for 5 minutes at 4000rpm in order to remove the liquid media and harvest the cells.

Mononucleosome isolation

We followed the protocol to isolate nucleosomes described by Shivaswamy et al. (Shivaswamy et al., 2008). Samples were treated with 250 μg of zymolyase (MP Biomedicals Catalog # IC320921) to permeabilize the cell wall. The cells were then washed and resuspended in NP buffer. The cells were then subjected to increasing concentrations of MNase (Worthington Biochemical Corp. Catalog # LS004797) at 25, 50, 75, and 100U/ml for 10 minutes at 37°C. Reactions were stopped by addition of 10mM EDTA and 1% SDS. Reverse crosslinking was performed by a 65°C overnight incubation with Proteinase K. RNA was then removed by RNase treatment. DNA was extracted by phenol chloroform treatment followed by ethanol precipitation. The DNA was then run on an E-gel cassette (from Invitrogen), and the fraction of DNA running at approximately 150bp was extracted.

Chromatin Immunoprecipitation

Yeast cells were fixed by adding formaldehyde to cultures at a concentration of 1% and incubating for 30 minutes at 30°C in a shaking incubator. Cells were then spun down, washed, and re-suspended in chilled lysis buffer and subjected to bead beating at 4°C. Samples were then sonicated using a Branson Sonifier, spun down, and the supernatant was isolated. A portion of the supernatant was reserved for an input sample and the remainder was used subjected to immunoprecipitation using either IgG Sepharose 6 Fast Flow beads (from GE Healthcare Life Sciences) or anti-Myc conjugated agarose beads (from Sigma Aldrich).

MNase Chromatin Immunoprecipitation

The MNase ChIP protocol was largely adapted from a published protocol (Wal and Pugh, 2012). Yeast cells were fixed by adding formaldehyde to cultures at a concentration of 1% and incubating for 30 minutes at 30°C in a shaking incubator. Cells were then spun down, washed, and re-suspended in chilled NP buffer and subjected to bead beating at 4°C. The cell lysate mixture was then collected and treated with MNase at increasing concentrations of 25, 50, 75, and 100U/ml for 10 minutes at 37°C. The reaction was then terminated by addition of EDTA to a final concentration of 10mM and incubating on ice for 10 minutes. The samples were then spun down at 4°C and the supernatants from the different MNase concentrations pooled and collected. A portion of the supernatant was reserved as an input sample, and the rest was used for chromatin immunoprecipitation. TAP-tagged proteins were pulled down by an overnight incubation at 4°C with IgG Sepharose 6 Fast Flow beads (from GE Healthcare Life Sciences).

Library preparation and sequencing

Libraries were prepared using the NEBNext ChIP-Seq library preparation kit for Illumina sequencing (NEB Catalog # E6240L) with adapters from Bioo. The libraries were then sequenced either at the University of Texas at Austin Genome Sequencing and Analysis Facility (UT GSAF) or at the M. D. Anderson Next-Generation Sequencing Facility at Science Park.

Analysis of sequencing data

FASTA files were aligned against the SacCer3 reference genome (from the Saccharomyces Genome Database) using the BWA alignment program. Wig files were then produced and uploaded into the UCSC Genome Browser. Peak files were also produced using an in lab peak calling pipeline. The peaks which are called are then used

to create a pergene file which consists of a window of 10bp bins centered either on the transcription start site (TSS) or the transcription termination site (TTS) or some other genomic loci of interest. These files are then normalized by in matrix normalization to allow comparison across files regardless of sequencing depth. This, in effect, is a normalization system centered around ORF regions, and bypasses having to take into account the large variability in nucleosome enrichment seen at repetitive regions. These normalized files can then be used to create an average TSS nucleosome profile, and nucleosome heatmaps based on the matrix can be visualized using JavaTree viewer. ChIP data was corrected by subtracting out the signal from matched input samples. For MNase ChIP-seq, input samples correspond to the background mono-nucleosome profiles (Figure 3.1).

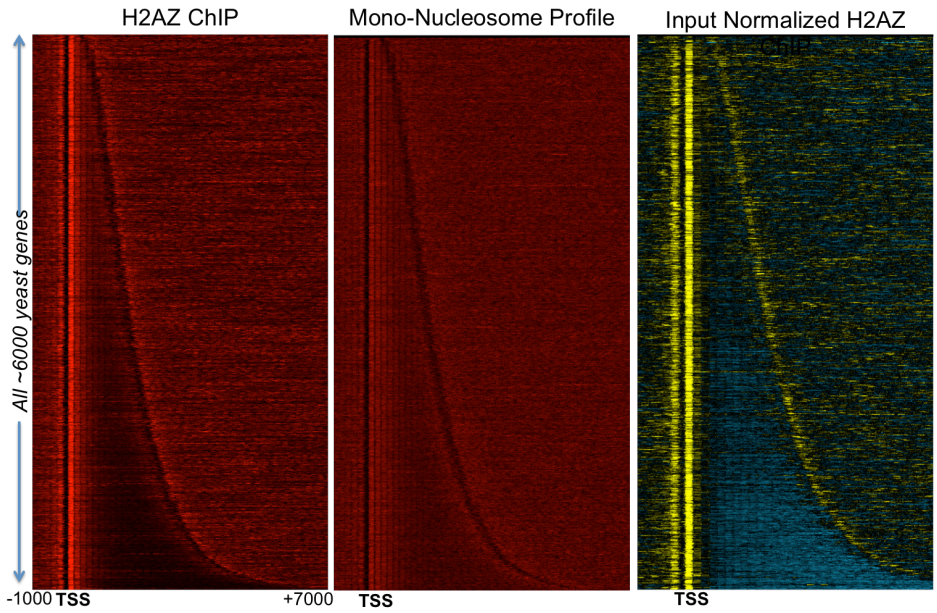

Figure 3.1: Uncorrected ChIP, mononucleosome, and input corrected ChIP heatmaps

In the middle background nucleosome profiles are depicted. On the left is a heatmap of H2A.Z enrichment without input correction. On the right is the input corrected ChIP

(Figure 3.1 continued.) data, with yellow indicating higher signal in the ChIP sample (enrichment) and blue indicating higher signal in the input sample (depletion). Transcripts are arranged by increasing transcript length, revealing a peak of H2A.Z at the 5' end of transcripts and a less pronounced peak at the 3' end of transcripts. The 3' peak is often associated with the 5' end of the nearest downstream gene.

Gene lists used for sorting

Gene expression data was obtained from publicly available RNA sequencing data derived from WT cells (van Dijk et al., 2011). Reads mapping to annotated transcripts were counted and then normalized for gene length. BNC counts were obtained from publicly available SMORE-seq data (Park et al., 2014a). Reads mapping between -50bp and -300bp upstream of the TSS for annotated genes were then counted to give a measure of upstream antisense transcription. For tandemly arranged transcripts, these antisense reads are referred to as BNCs. TATA and TATA-less gene lists were obtained from Rhee et al. (Rhee and Pugh, 2012). A ribosomal protein coding gene list was obtained from The Ribosomal Protein Gene Database (Nakao et al., 2004). Tandem and divergent gene orientations were determined computational by annotating each TSS with whether the nearest upstream gene end was a transcription start or termination site. H2A.Z enrichment values for +1 and -1 nucleosomes were determined by taking the maximum signal across 150bp within a specified range (Figure 3.2).

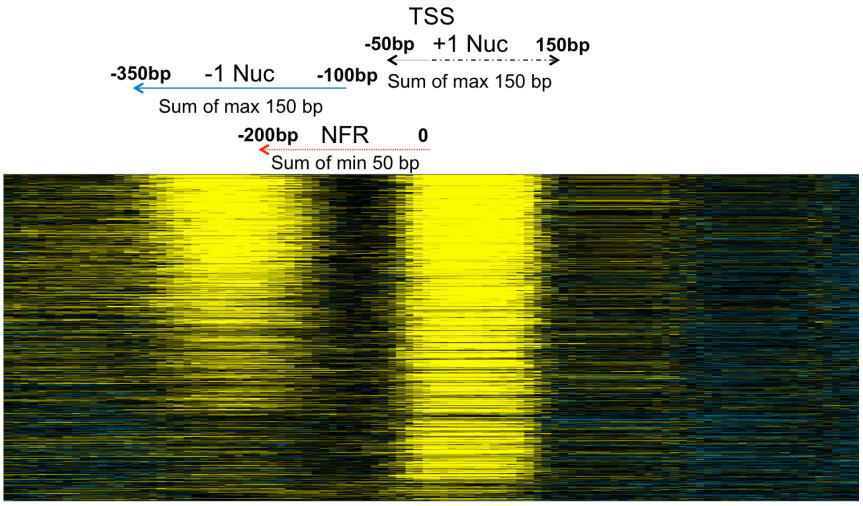

Figure 3.2: Schematic depicting the method used to call enrichment values for the -1 and +1 nucleosome

Above is a heatmap detailing the ranges over which the +1 nucleosome, NDR, and -1 nucleosomes predominantly fall. Within these windows a +1 nuc, NDR, and -1 nuc value was called for each gene. For nucleosomes this value corresponds to the maximum signal within 15 10bp bins (for 150bp) across the window. For NDRs the value corresponds to the minimum signal within 5 10bp bins (for 50bps) within the NDR window.

RESULTS

In order to explore H2A.Z incorporation at nucleosomes across the *S. cerevisiae* genome we performed MNase ChIP experiments to pull down TAP-tagged H2A.Z. As a control, input samples were also produced. This allowed us to determine nucleosomes within the genome at which H2A.Z is incorporated as well as nucleosomes at which it is not. The results of two replicates for the MNase ChIP were fairly consistent, so the average of the two was used for further analysis (Figure 3.3). As the chromatin remodelers Swr1 and Ino80 govern the incorporation dynamics of H2A.Z, we also performed ChIP experiments for Swr1 and Ino80 Myc tagged strains in both WT and in

ΔHtz1 backgrounds. An additional Htz1-TAP ChIP was also performed for comparison with the MNase-ChIP.

Analysis of the ChIP data revealed Swr1 and Ino80 binding to be fairly low and consistent across most genes, with only a handful of genes showing strong binding peaks. What was especially striking was that Swr1's strongest peak (an extreme outlier in our data) was localized at its own promoter (Figure 3.4). This result was confirmed by examining publicly available Swr1 ChIP data produced by the Pugh lab (Yen, 2013, 24034248) and similar observations from microarray data have been made before (Zhang et al., 2005). That Swr1 should bind its own promoter so strongly, was surprising, and prompted us to look at H2A.Z incorporation at the promoter as well. ChIP data for H2A.Z revealed a pronounced peak at the upstream gene, but rather weak enrichment at the Swr1 binding site (Figure 3.4). Furthermore, we found that deletion of Htz1 decreased the Swr1 binding peak height compared to the background and led to a pronounced change in shape, with Swr1 spreading farther upstream into the region where the H2A.Z peak is found in WT cells. This binding pattern suggests a feedback system that might regulate Swr1 transcription based on Swr1 levels in the cell.

Overall, the Swr1 ChIP data displays a weak positive correlation with H2A.Z localization (Figure 3.5B). Given the large dynamic range present in the data, it seems unlikely that the variance in Swr1 localization alone can account for the variance seen in H2A.Z deposition. It is interesting to note that Swr1 has been reported to switch to a "promiscuous" mode in the context of acetylation of H3 at lysine 56 (H3K56Ac) where it can exchange either H2A or H2A.Z containing dimers (Watanabe et al., 2013). There is also some evidence for random incorporation of H2A.Z, which is not Swr1-dependent (Hardy et al., 2009, Hardy and Robert, 2010). Likewise, genes that require H2A.Z for

expression and not Swr1, as well as genes that require Swr1 but not H2A.Z have been reported. Any, and all of these factors may contribution to the low correlations.

To account for differences that might exist between two divergent (head-to-head) transcripts and two tandem (tail-to-head) ones, genes were separated based on their TSS orientation. They were then grouped by their Swr1 enrichment (top 500, bottom 500, and middle for all other genes) and their average H2A.Z profiles were plotted. While tandem and divergent genes showed very similar levels of Swr1 localization across the ranges (data not shown), their enrichment patterns for H2A.Z differed (Figure 3.5A). Both had roughly equivalent + 1nucleosome enrichment levels, however divergent genes had noticeably more H2A.Z at their -1 nucleosome. Additionally, the tandem gene -1 nucleosome was primarily a single peak and was not associated with additional regularly spaced arrays of H2A.Z containing nucleosomes decaying upstream.

We next explored how Swr1 and Ino80 binding patterns were distributed across the genome. First genes were sorted by ChIP binding levels (Figure 3.6 A and C). Next, average binding profiles were produced for the top 500 most bound and bottom 500 least bound targets. Profiles for all other genes are labeled as the middle group. These plots revealed that at a handful of targets, Swr1 and Ino80 could be found at the +1 nucleosome, while at the majority of genes binding levels did not exceed the values seen in the inputs (Figure 3.6 B and D). The data also revealed that these high binding targets were redistributed in the absence of H2A.Z. This redistribution was more profound for Ino80 as both the least and most enriched sites changed. By contrast, sites with minimal Swr1 binding in the WT remained depleted in the ΔHtz1 strain. Since Ino80 recognizes and evicts H2A.Z/H2B dimers (found in a fraction of the total nucleosomes), its more pronounced redistribution in the absence of the H2A.Z histone is somewhat understandable.

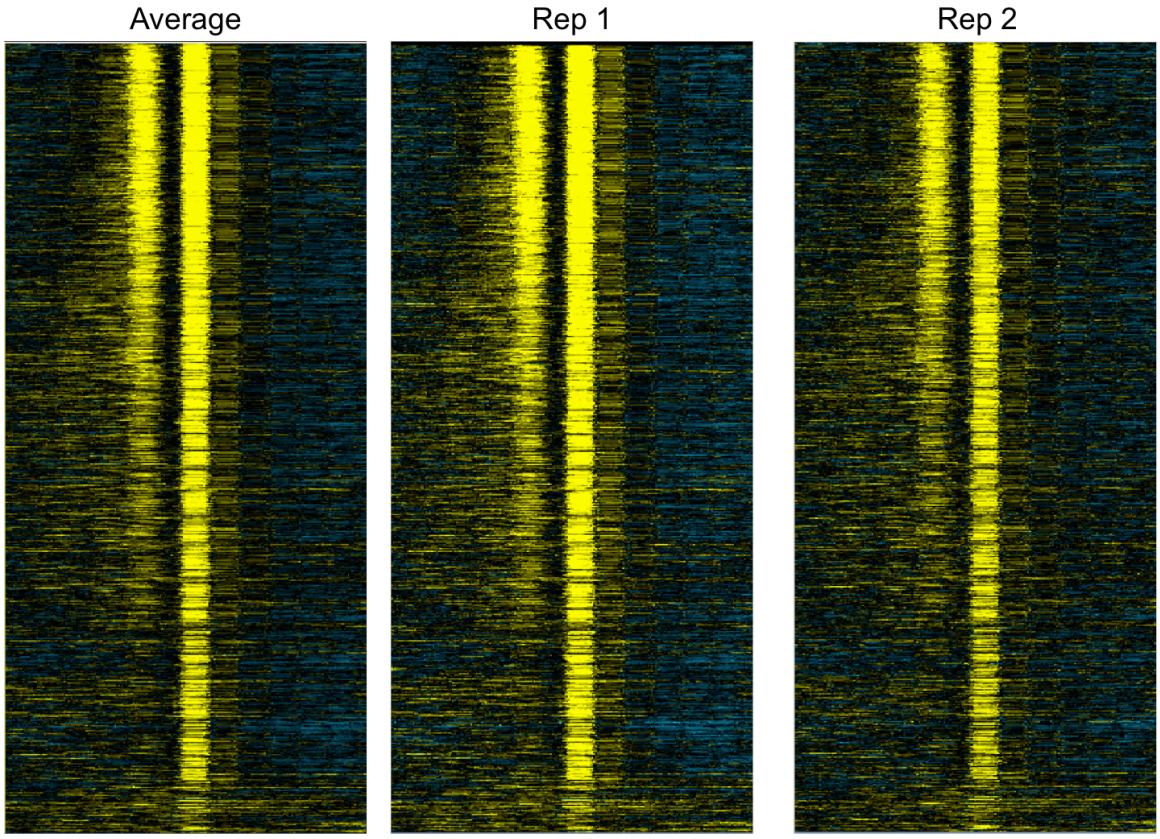

Figure 3.3: H2A.Z MNase ChIP replicates and averaged data

Displayed above are heatmap representations of Input normalized ChIP seq. The two replicates showed very similar enrichment patterns, so, the average of the two replicates was taken and was used for further analyses.

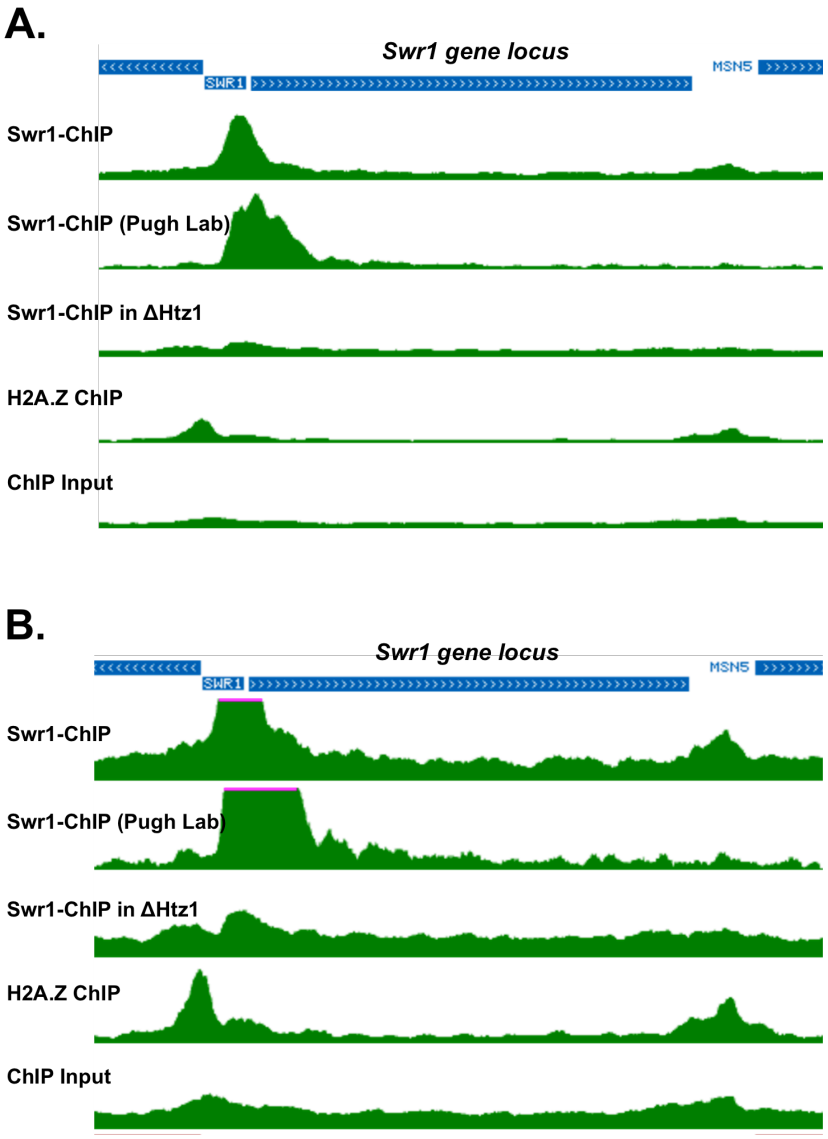

Figure 3.4: Swr1 binds its own promoter

The data displayed above has been visualized with a mirror of the UCSC Genome Browser. B displays a more zoomed in version of A, and in both the backgrounds have been equalized between tracks. The top two rows contain Swr1-ChIP binding data (the 1st contains data from our lab and the 2nd data downloaded which was produced by the Pugh Lab). The 3rd row contains Swr1-ChIP data in a ΔHtz1 strain. The ChIP signal within this mutant is greatly reduced (this can be seen from the increased background compared to signal in this track), and partially shifted upstream, compared to the wildtype strain. The 4th row contains H2A.Z ChIP-seq data, and the last row input data for comparison.

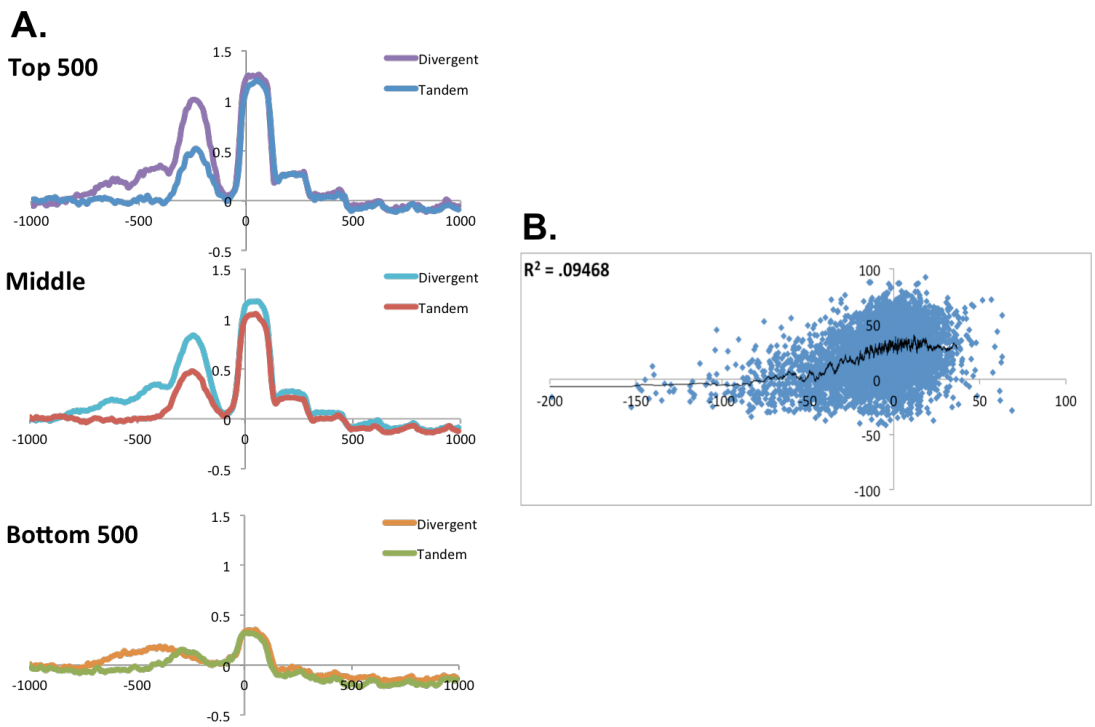

Figure 3.5: H2A.Z localization in response to Swr1 localization

(A) Plotted are H2A.Z enrichment levels from -1000 to +1000 around the TSS. Gene groups have been determined based on the intensity of input corrected Swr1 binding at the NDR, as measured by the total signal between -350 to +200 (this window was picked to try to encompass all NDR associated binding, see Figure 3.2 for region estimates). The plots further separate tandem from divergent genes. Consistently, tandem genes are associated with a single upstream peak of H2A.Z, which, regardless of Swr1 binding level, is much lower than that see at divergent genes. (B) A plot of the correlation between Swr1 ChIP signal between -350 to +200 and the H2A.Z signal over the same region. The correlation is rather low at ~.095. A handful of Swr1 input corrected enrichment values below -200 and above +100 have been dropped to allow plotting, but were included in correlation estimates. The black line represents a moving average across 50 genes at a time.

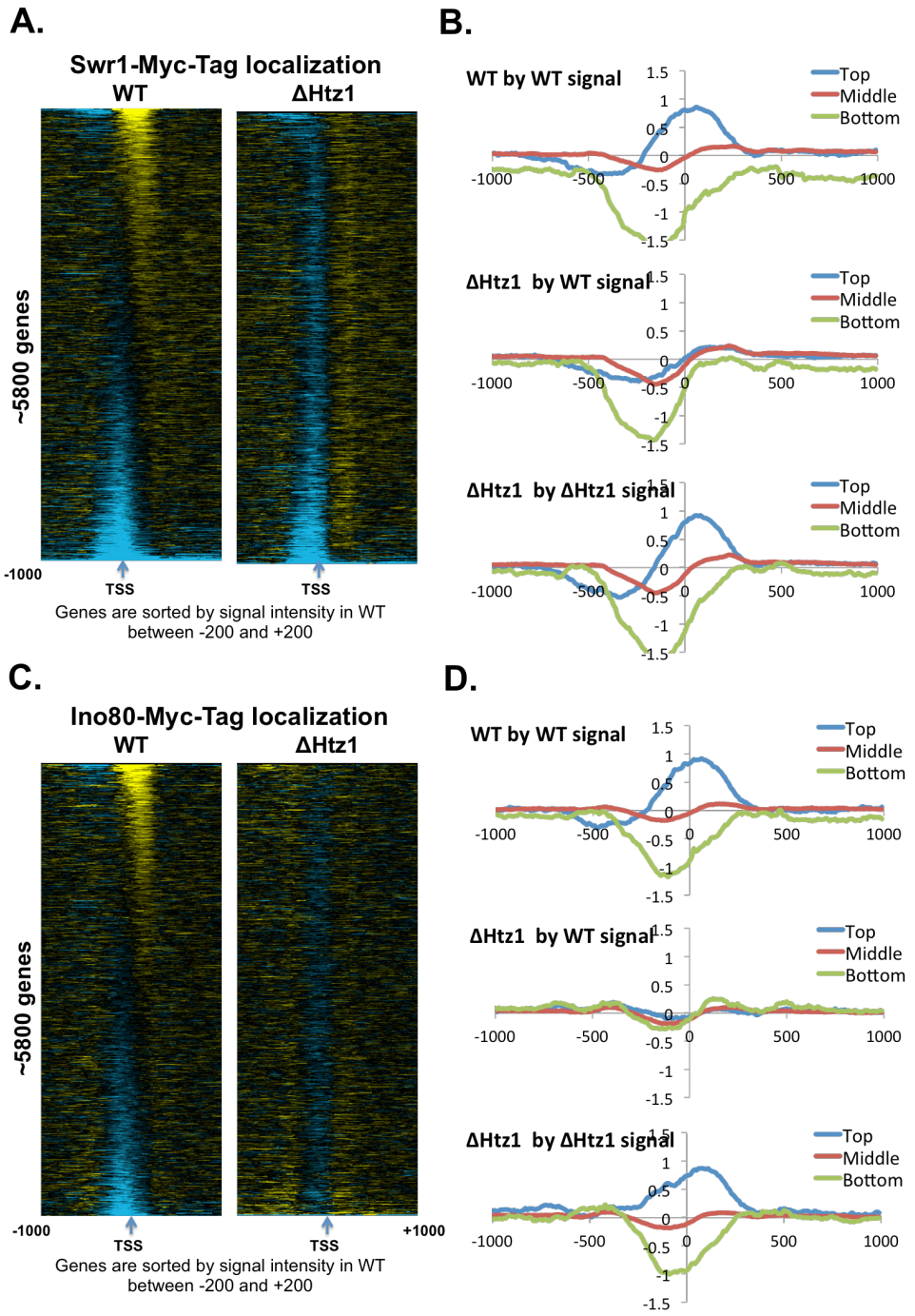

Figure 3.6: Swr1 and Ino80 binding targets are redistributed in the Htz1 deletion strain

(A and C) Heatmaps of Swr1 and Ino80 ChIP data in WT and ΔHtz1 strains sorted by the signal in the WT data between -200 and +200 bp surrounding the TSS. (B and D)

(Figure 3.6 continued.) Average binding profiles of Swr1 and Ino80 across the TSS. Gene groups are determined as follows: Top- includes data from the top 500 genes by signal, Bottom –includes data from the 500 least genes by signal, Middle –includes data for the remaining 4797 annotated genes. The top plot for each contains WT data. The next two plots contain data from the ΔHtz1 strains, 1st with the groups determined by the signal in the WT sample, and 2nd with the groups determined by the signal in the mutant sample. These plots indicate that the main strong binding locations within the genome of Swr1 and Ino80 are redistributed in the context of loss of H2A.Z.

H2A.Z and gene expression

Next we decided to look at the impact of gene expression on genome wide H2A.Z patterns. Previously, data from microarray experiments indicated that H2A.Z might be inhibitory to transcriptional initiation, as occupancy was shown to correlate negatively with transcription rate (Li et al., 2005, Zhang et al., 2005). However, when H2A.Z is acetylated it is associated with actively transcribed genes (Millar et al., 2006). There is also evidence that H2A.Z promotes transcriptional elongation. In its absence nucleosome occupancy increases over the GAL10p-VPS13 gene locus, and the elongation rate of RNA Pol II decreases by ~24% (Santisteban et al., 2011). Additionally, phosphorylation of RNA Pol II's Ser2 residue within the gene body is deficient.

We began by examining whether there was a correlation between H2A.Z enrichment levels at the +1 nucleosome and gene expression levels in our data. Sorting transcripts based on their expression level failed to produce an overall linear correlation with H2A.Z enrichment (Figure 3.7). Instead we found that genes with very high expression or very low expression both showed low levels of H2A.Z enrichment, while the majority of genes displayed fairly uniform levels of H2A.Z. In the case of highly expressed genes, the depletion of H2A.Z at the +1 nucleosome could reflect higher transcription rates displacing H2A.Z too quickly for detection by MNase-ChIP.

Analysis of Swr1 and Ino80 binding patterns at highly expressed genes revealed that, under normal conditions, these proteins also display reduced binding at these locations (Figure 3.8). However, in the absence of H2A.Z, Ino80 accumulated upstream of these genes and within the NDR. This increase may suggest that Ino80 is normally targeted to the upstream region to regulate the displacement of +1 nucleosomes at these genes. However, in the absence of H2A.Z, it is without a substrate to act upon, and it accumulates at the NDR.

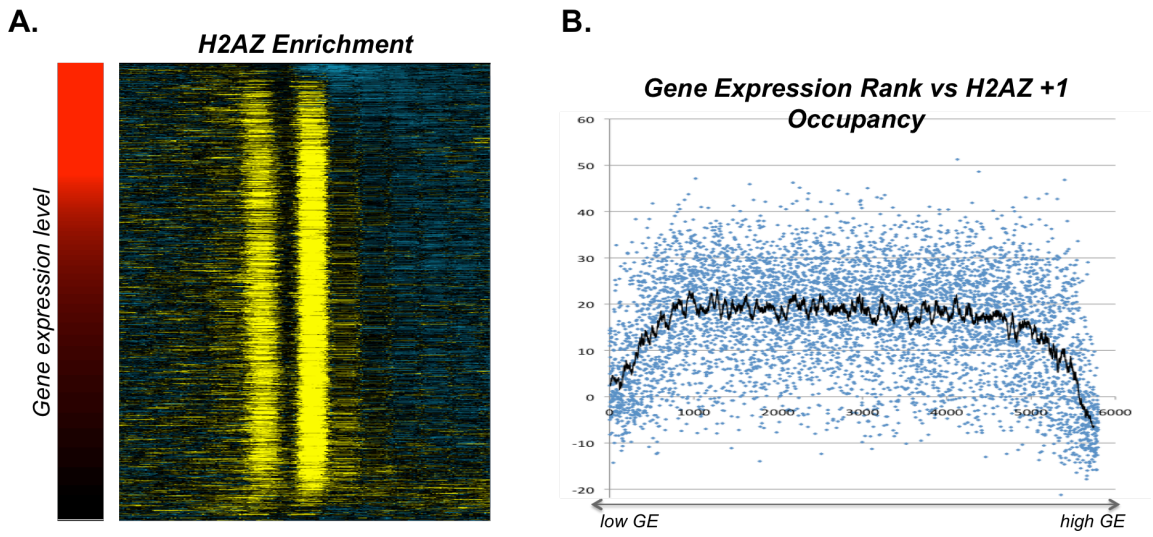

Figure 3.7: H2A.Z enrichment at the +1 nucleosome does not correlate with gene expression

(A) A heatmap (visualized with Java Tree Viewer) displaying the H2A.Z enrichment across transcript TSS regions when sorted by gene expression level. The x axis represents distance from the TSS within a range of -1000bp upstream to +1000bp downstream. (B) A plot of H2A.Z occupancy at the +1 nucleosome vs a ranked gene expression value. Ranking was used to accommodate outliers that skewed the plot. The black line represents a 50 gene moving average of H2A.Z enrichment.

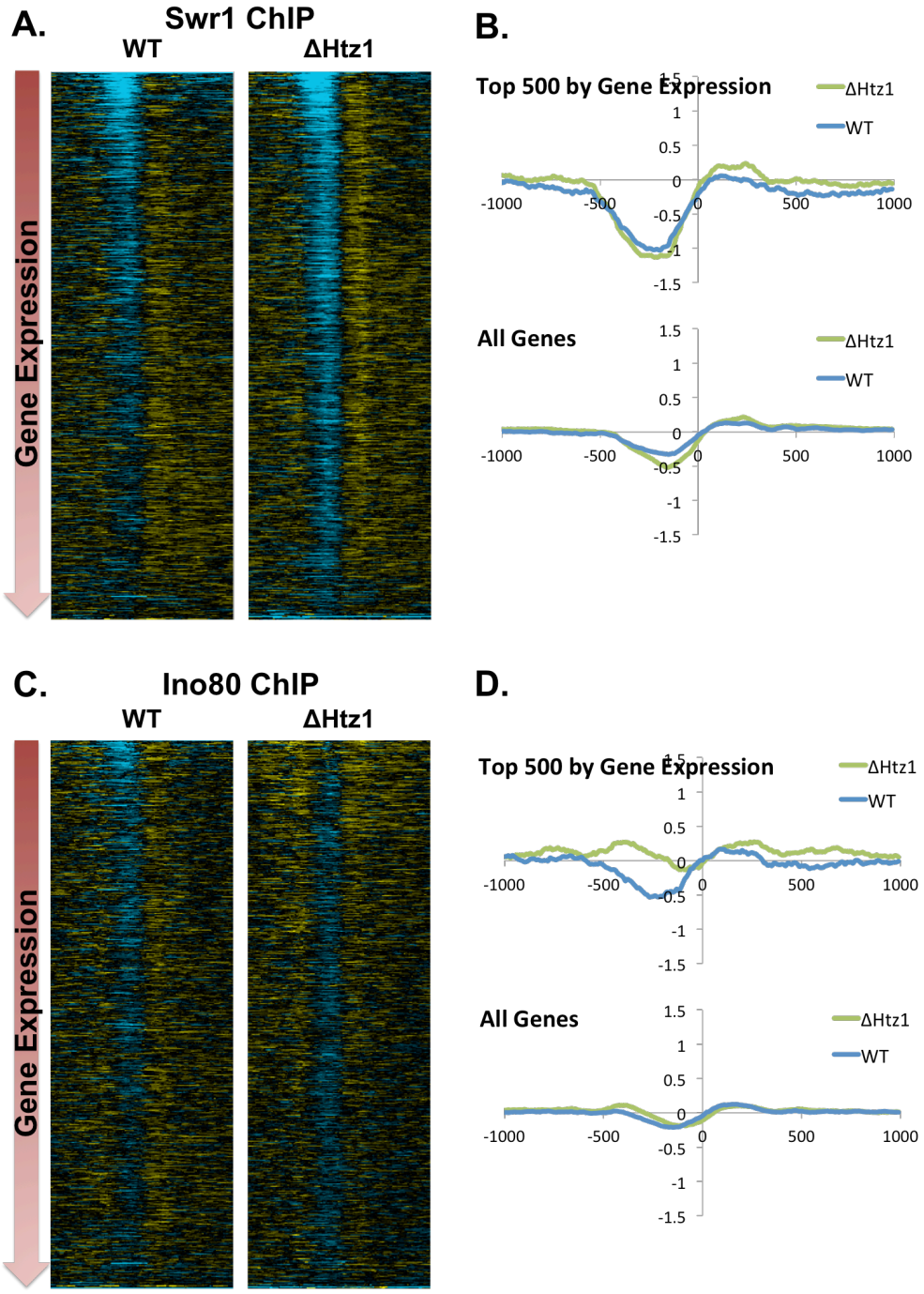

Figure 3.8: Swr1 and Ino80 localization at highly expressed genes in an Htz1 deletion strain

(Figure 3.8 continued.) (A) Heatmap of Swr1 ChIP data in WT and ΔHtz1 strains sorted by gene expression level. (B) Average binding profiles of Swr1 across the TSS. Top is the signal for the 500 most highly expressed genes, while at bottom is the signal across all yeast genes. (C) Heatmap of Ino80 ChIP data in WT and ΔHtz1 strains sorted by gene expression level. (D) Average binding profiles of Ino80 across the TSS for the 500 most highly expressed genes and for all yeast genes.

H2A.Z enrichment at the +1 nucleosome correlates with upstream antisense transcription

A number of studies have described high levels of antisense transcription occurring upstream of gene promoters (Neil et al., 2009). As this antisense transcription might also contribute to the chromatin structure found at the TSS, we were interested in investigating whether the H2A.Z signal was influenced by upstream antisense transcription. H2A.Z has already been reported to exhibit co-occupancy with antisense transcription emanating from the 3' ends of transcripts (Gu et al., 2015). For this analysis we used antisense transcription measurements from publicly available SMORE-seq data (Park et al., 2013), and in particular their calls for antisense transcription upstream of tandem genes, which are referred to as bidirectional noncoding RNAs or (BNCs).

When we separated genes by TSS orientation, and aligned them by upstream transcription level within a window designed to allow detection of BNCs (from -300 to -50 upstream of the TSS), we noted that both groups appeared to show a correlation between antisense transcription and H2A.Z occupancy at the -1 nucleosome (Figure 3.9). However, the divergent genes displayed significant amounts of H2A.Z enrichment in the nucleosomal arrays extending further upstream from the NDR. Aligning these genes by distance to the upstream TSS revealed that much of the diffuse signal was a result of upstream TSSs that were farther away. It also revealed that genes with the highest H2A.Z enrichment levels correspond to ones where two diverging coding genes share an

NDR, and presumably a bidirectional promoter. For the remainder of genes, there was a notable pattern of stronger antisense transcription corresponding to stronger H2A.Z signal at the -1 nucleosome. We investigated this association by separating genes based on BNC levels, and determining whether there were significant differences in H2A.Z enrichment among the groups (Figure 3.11). All comparisons yielded significant p-values.

We also determined that levels of Swr1 enrichment were roughly equivalent between tandem and divergent genes (data not shown), suggesting that increased Swr1 targeting to the NDR does not explain the increased H2A.Z incorporation at divergent genes. It became apparent that divergent transcripts had some H2A.Z incorporated at the -2 and-3 nucleosomes, whereas the tandem genes only displayed this enrichment at the -1 (Figure 3.12). This pattern is consistent with increased displacement occurring at divergent genes where upstream transcripts will normally be long and coding. By contrast, antisense upstream reads associated with tandem gene promoters are short, and transcription at these regions could, therefore, be less likely to displace H2A.Z in the direction of transcription.

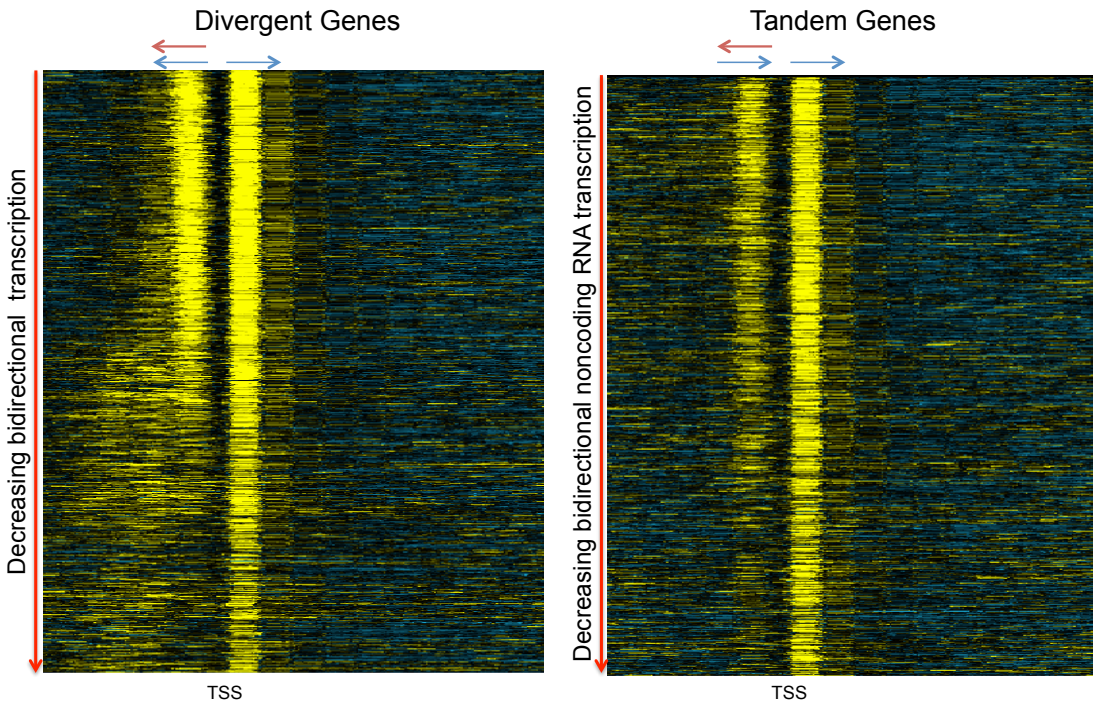

Figure 3.9: Aligning H2A.Z enrichment data by upstream antisense transcription level

H2A.Z enrichment data has been segregated based on the orientation of the upstream gene at the TSS. Divergent genes are characterized by a head to head orientation, while tandem genes are transcribed in the same direction and on the same strand resulting in a tail to head orientation. Gene tails correspond to arrow tips in the schematics displayed above the heatmaps. The data has been further arranged by antisense transcription level in decreasing order. The x-axis includes a range around each transcript's annotated TSS of -1000 to +2000bp.

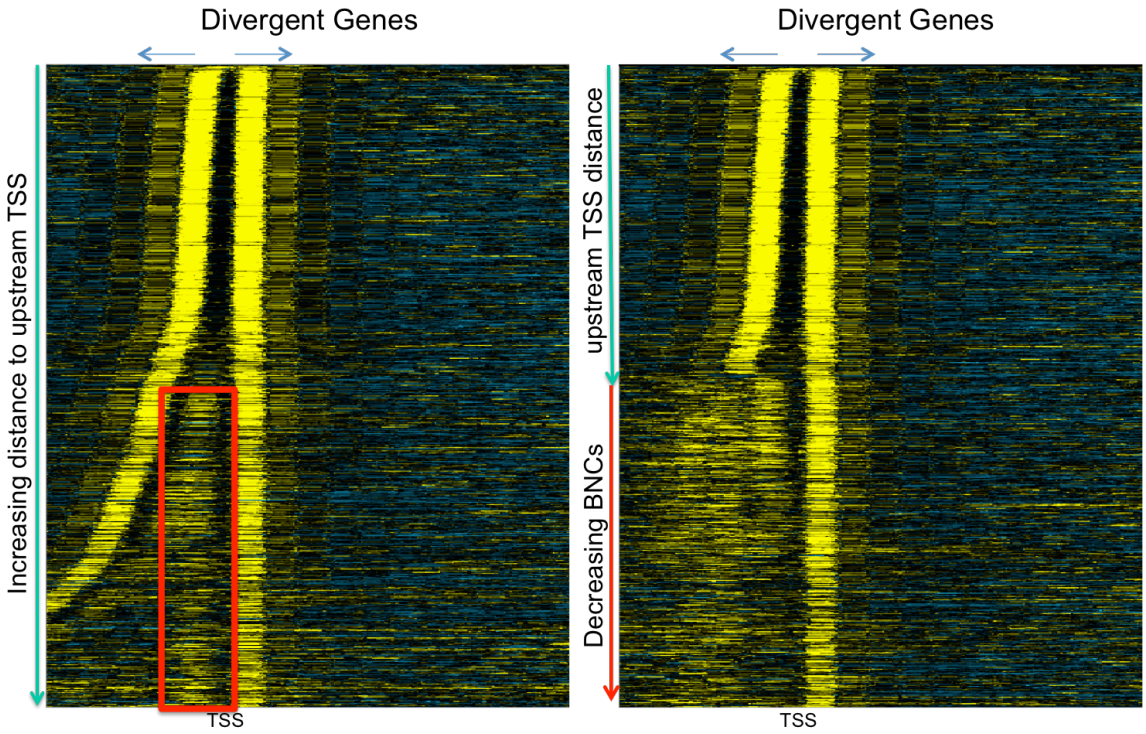

Figure 3.10: Detecting noncoding antisense reads between divergent transcripts

The apparent correlation between antisense transcription and H2A.Z incorporation at the -1 nucleosome at divergent genes reflects differences in the intergenic distance between transcript TSSs. On the left, genes are sorted based on the distance to the upstream TSS. On the right, genes with a TSS-to-TSS distance greater than 300bps are resorted based on the level of antisense transcription, revealing a pattern of H2A.Z enrichment matching the strength of BNCs emanating from these NDRs.

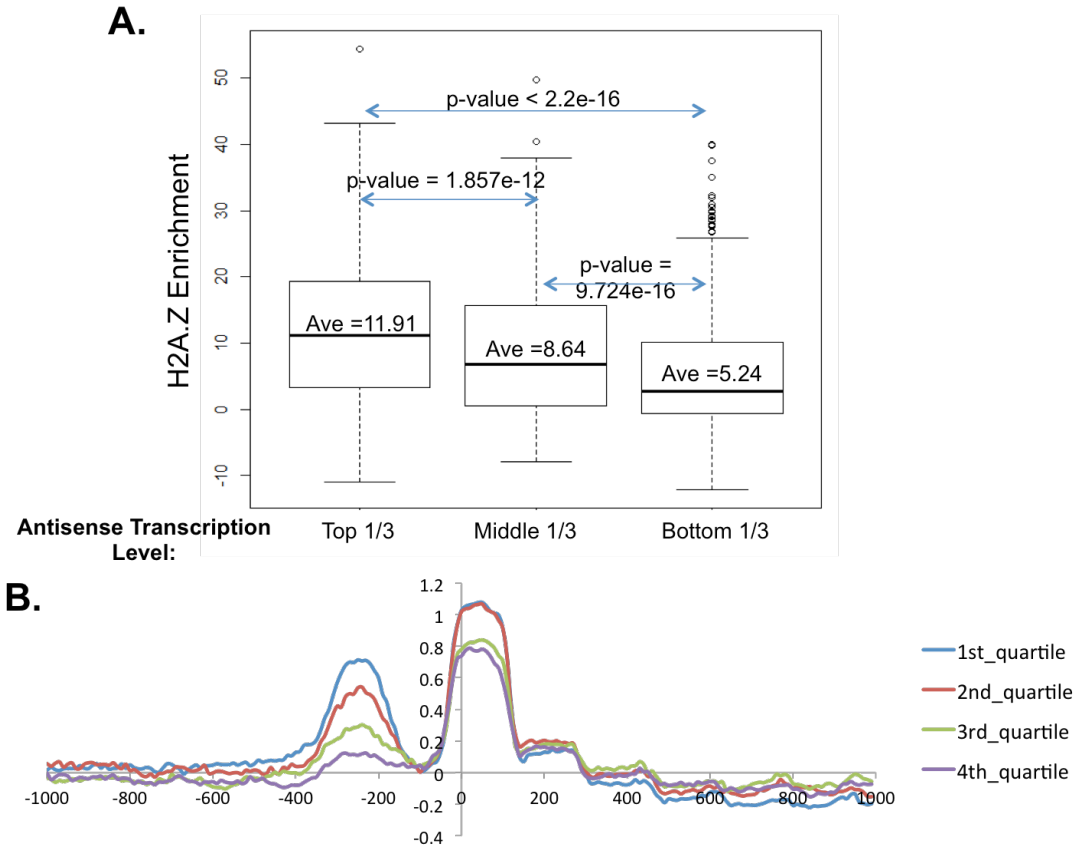

Figure 3.11: H2A.Z enrichment at the -1 nucleosome of tandem transcripts increases with increasing antisense transcription

In both images, tandem transcripts are grouped according to antisense transcription level. (A) Boxplots of H2A.Z enrichment at the -1 nucleosome. Welch t-tests were performed to compare the averages between the 3 groups, yielding p-values indicating that the differences were significant. (B) Average profiles are plotted for the transcripts separated into quartiles. The plot clearly shows that, overall, H2A.Z enrichment at the -1 nucleosome increases in concert with increasing antisense transcription level.

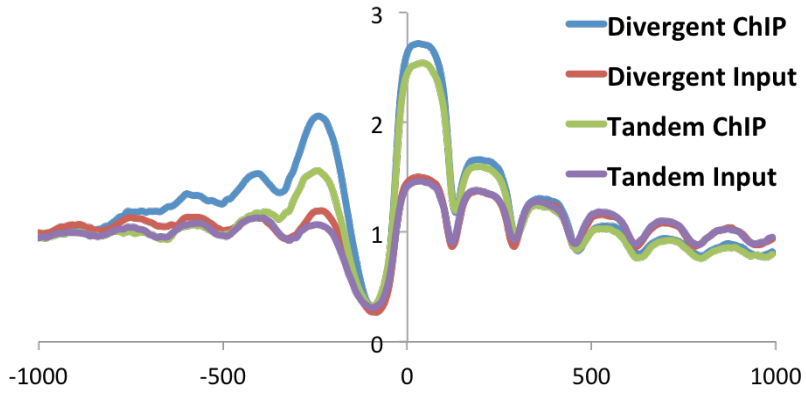

Figure 3.12: Average profile of input vs. non-input corrected H2A.Z MNase ChIP by TSS orientation

H2A.Z incorporation at the NDR differs dramatically between divergent and tandem transcripts. By contrast, background nucleosome levels (input) show little difference between the two groups.

The TATA box

In general, genes with TATA box containing promoters exhibit different transcriptional characteristics. On average, they produce higher levels of gene expression, and much lower levels of upstream antisense transcription (Figure 3.13). These features are in agreement with the observation that TATA boxes are generally enriched at sites of strong sense directed transcription. Genes with TATA containing promoters also exhibit more dynamic expression levels and are enriched for up-regulation during environmental stress and depleted for housekeeping genes (Basehoar et al., 2004). Since H2A.Z marks sites of bidirectional transcription, we wanted to explore H2A.Z incorporation at TATA box associated TSSs. Zhang et. al previously demonstrated by microarray techniques that, when using a sliding window of 80 genes, regions of the genome with higher numbers of TATA boxes displayed less H2A.Z occupancy (Zhang et al., 2005). Modern sequencing methods have increased the resolution that can be

obtained through such studies. Consequently, we decided to investigate the impact of TATA boxes on H2A.Z incorporation at the TSS on a genome wide scale.

We plotted average H2A.Z enrichment levels for TATA possessing and TATA-less genes, and visualized the gene-by-gene enrichment by producing a heatmap of enrichment with JavaTree Viewer (Figure 3.14). The heatmap revealed that H2A.Z does localize to the TSS of TATA possessing genes. However, the enrichment is lower and the signal more diffuse than the precise and strong signal witnessed at TATA-less genes. The average profile reveals that, at TATA containing genes, less H2A.Z is incorporated into the +1 and -1 nucleosomes, and that H2A.Z can be found at low levels throughout the gene body and within the NDR. These results indicate that, when a TATA box is present, H2A.Z is less likely to be incorporated in nucleosomes surrounding the NDR, but also that when it is incorporated the localization is not as strictly maintained. This may reflect higher rates of transcription at TATA containing genes displacing H2A.Z containing nucleosomes farther down the gene body. Since the Swr1 and Ino80 complexes share a number of subunits, the profiles could also indicate, that, regions which are less likely to be targeted for H2A.Z incorporation by Swr1 are also less likely to be targeted for H2A.Z removal by Ino80. A t-test based comparison between H2A.Z levels at +1 nucleosomes for TATA containing vs. TATA-less genes yielded a p-value of $< 2.2e-16$ (Figure 3.15). Analysis of the differences between Swr1 and Ino80 binding patterns at these two gene groups also revealed that Ino80 accumulated at the TATA containing genes in the absence of H2A.Z, while in WT cells both complexes were depleted from TATA containing genes (Figure 3.16). These patterns were similar to its accumulation at highly expressed genes. They also suggest that low levels of H2A.Z at TATA containing genes are actively maintained by the actions of Ino80.

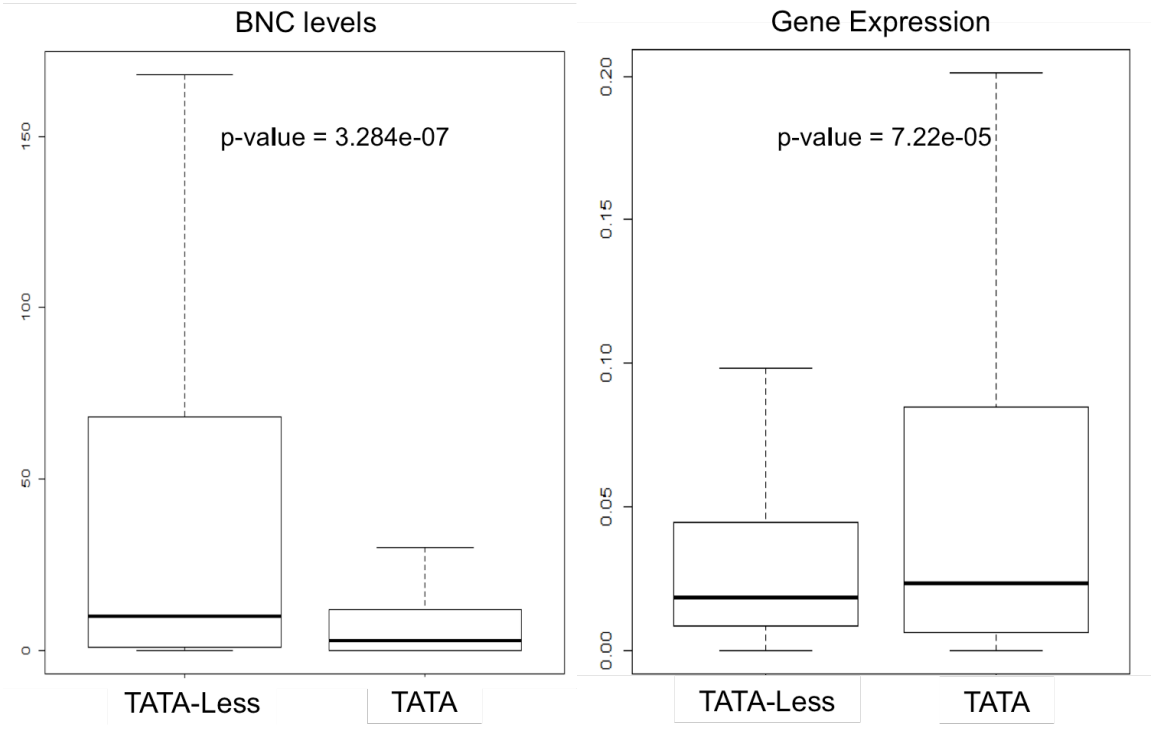

Figure 3.13: Transcriptional characteristics of TATA-box possessing genes

Displayed above are box plots for antisense transcription (BNCs) and gene expression levels. Outliers have been dropped. TATA containing and TATA-less genes show notable differences in the transcription emanating from their NDRs. TATA genes have higher expression levels overall, and are decidedly less likely to be associated with upstream antisense transcription. By contrast TATA-less genes display more modest levels of transcription downstream, but also increased levels upstream. Welch t-tests were performed for significance.

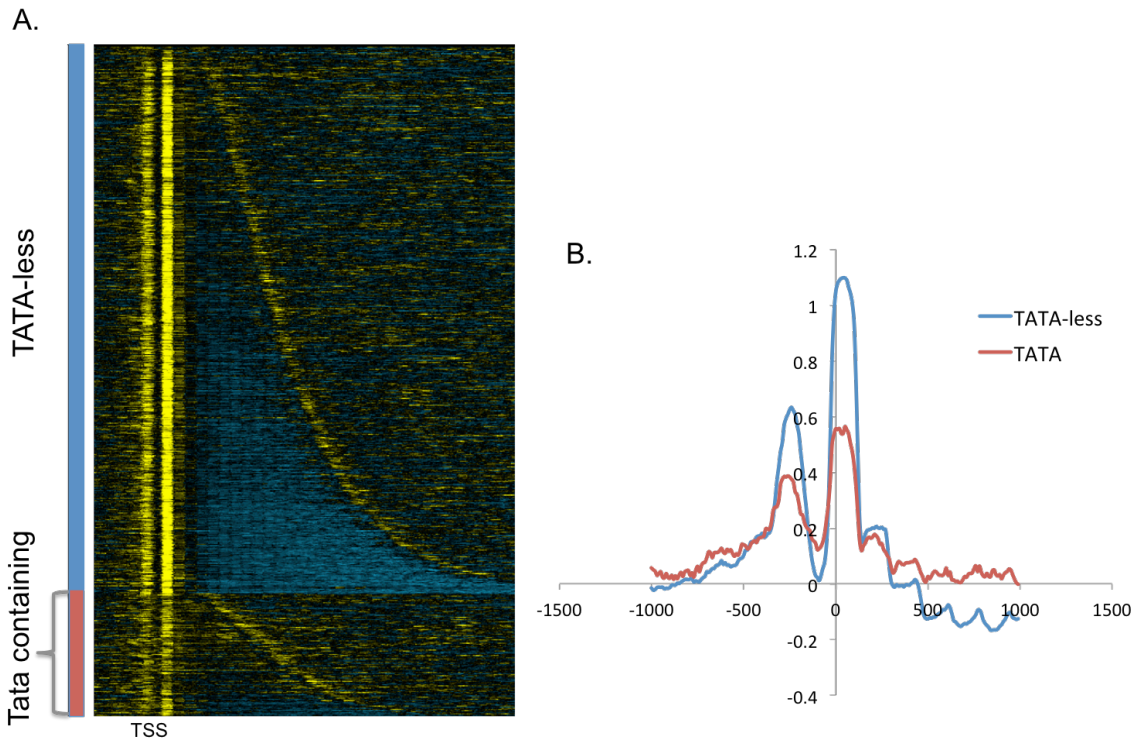

Figure 3.14: H2A.Z localization at the TSS with respect to the occurrence or absence of a TATA box

(A) A heatmap of H2A.Z localization separated by TATA promoter status. (B) Average H2A.Z enrichment levels across the two gene groups. Genes with TATA boxes show less +1 and -1 nucleosome incorporation of H2A.Z, but also display increased H2A.Z levels in the gene body.

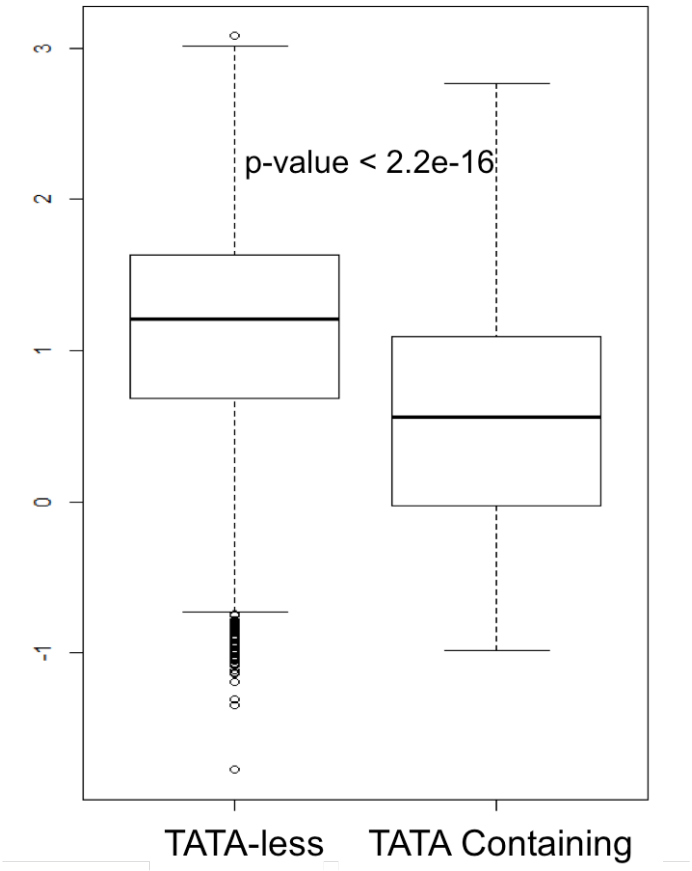

Figure 3.15: Boxplots of H2A.Z levels at +1 nucleosomes of TATA box containing and TATA-less promoters

A Welch t-test was performed for significance.

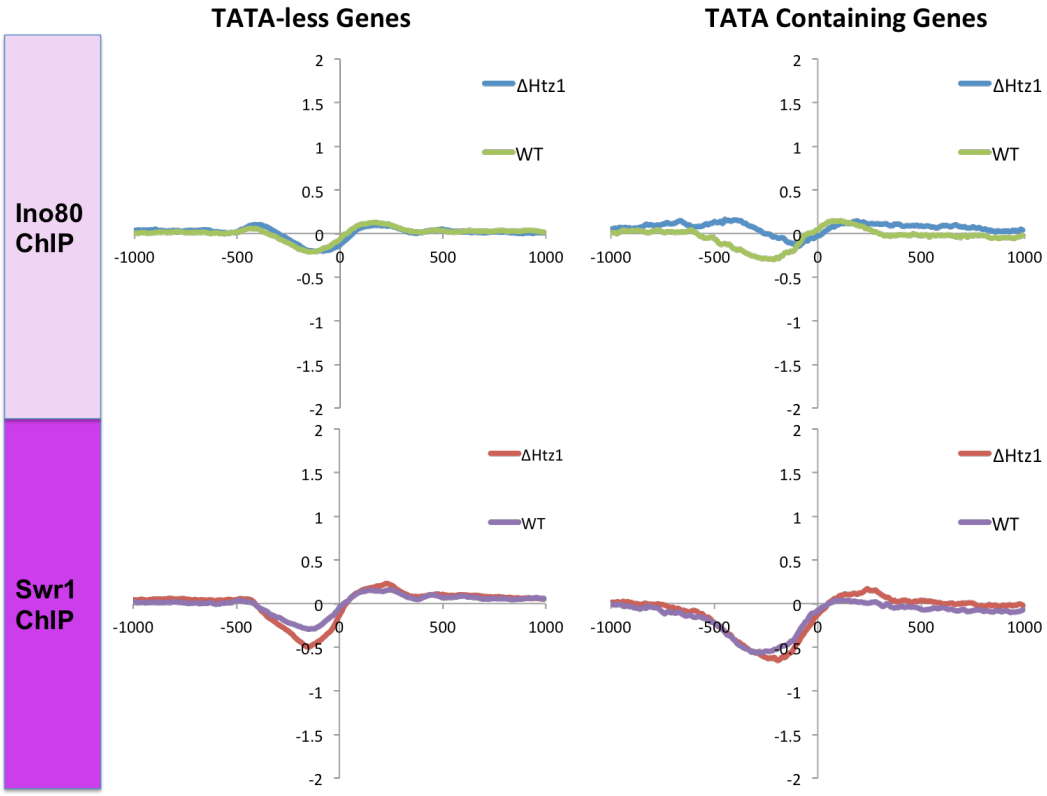

Figure 3.16: Ino80 and Swr1 localization at the TSS based on TATA status

Overall, Swr1 levels are low at TATA genes, especially in the NDR. Ino80 levels at TATA genes are low in the presence of histone H2A.Z, but show a modest increase in the upstream direction in its absence.

Ribosomal protein coding genes

Finally, we turned our attention to the highly expressed ribosomal protein coding genes. The 78 proteins in the yeast ribosome are encoded for by 137 gene loci (Warner, 1999). These genes were previously shown to be depleted for H2A.Z, in contrast to the enrichments seen at mitochondrial RP genes and ribosome biogenesis genes (Zhang et al., 2005). We also found a pronounced depletion of H2A.Z within these genes (Figure 3.17). Swr1 was, extremely depleted from the NDRs associated with these genes, but did exhibit of a modest binding peak at the +1 nucleosome. However, Ino80 was once again

detected at the NDR in the absence of H2A.Z. In light of previous evidence that some random incorporation of H2A.Z in the absence of Swr1 may occur within the genome (Hardy et al., 2009, Hardy and Robert, 2010), Ino80's association with RP genes may indicate that preventing the accumulation of H2A.Z containing nucleosomes within the NDR of these genes is an active process.

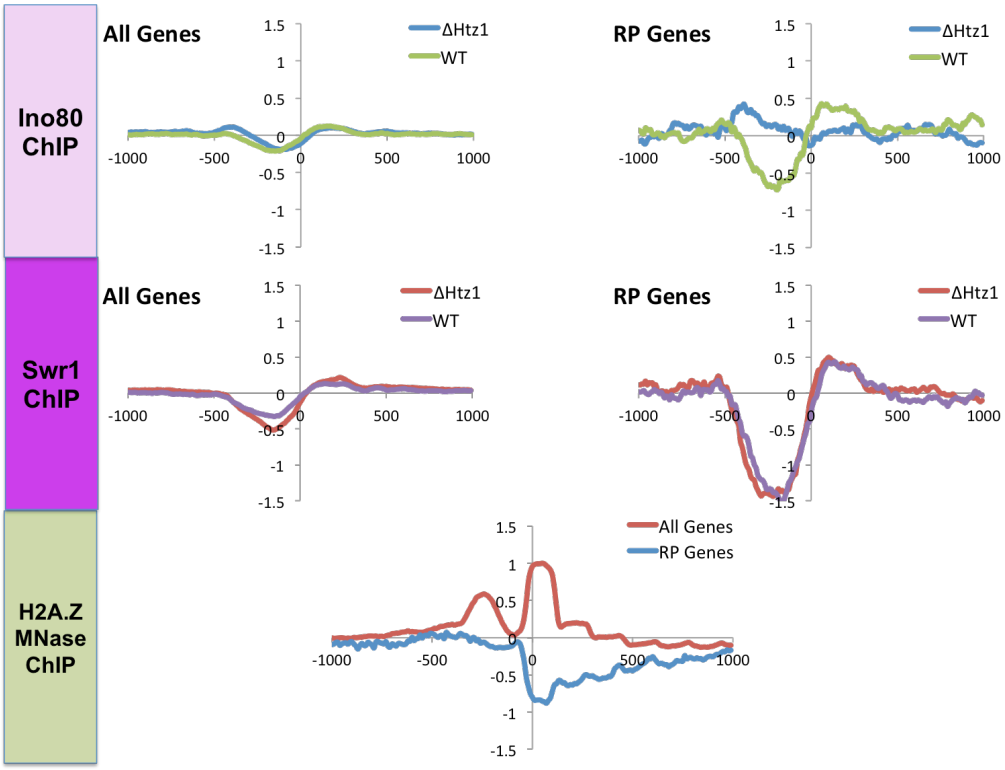

Figure 3.17: Ino80, Swr1, and H2A.Z dynamics at ribosomal protein genes

DISCUSSION

Our finding, that H2A.Z localization at both sides of the NDR is associated with upstream transcription, argues against a +1, NDR, -1 model, and instead, contends that NDRs with upstream and downstream H2A.Z are encompassed by two +1 nucleosomes (Figure 3.18). In light of our results, it is interesting to note, that, the NDR structure has been characterized differently in different organisms. In humans, peaks of H2A.Z enrichment are also found at both sides of the NDR, and correspondingly, bidirectional promoters producing bidirectional transcription are common (Preker et al., 2011, Trinklein et al., 2004). In contrast, the Drosophila NDR shows H2A.Z enrichment at the +1 nucleosome but not at the -1 (Mavrich et al., 2008). It is enticing to speculate, that, the lack of an upstream -1 nucleosome containing H2AZ could result from the fact that Drosophila gene promoters also display a pronounced lack of bidirectional transcription and a larger number of directional motifs (Core et al., 2012). Adding further weight to the importance of the +1, *Schizosaccharomyces pombe* predominantly lack -1 nucleosomes and regular nucleosomal arrays upstream of the NDR, while their downstream arrays also appear to form co-directionally with transcription (Lantermann et al., 2010). It has been noted, however, that a small subset of genes produce weak upstream nucleosomal arrays and that these promoters often contain H2A.Z. The formation of nucleosomal arrays co-transcriptionally, is in agreement with our data, as, BNC associated tandem promoters produce single peaks of H2A.Z upstream and not arrays, commensurate with the short transcript lengths of BNCs.

At first glance, the reports about H2A.Z's association with antisense transcription in *S. pombe* and *S. cerevisiae* seem contradictory, with one claiming H2A.Z suppressed antisense transcription, and the other claiming it had positive effects on it (Gu et al., 2015, Zofall et al., 2009). The *S. pombe* study looked specifically at long anti-sense

transcripts that are primarily produced from 3' originating convergent transcription. In this case, 5' incorporation of H2A.Z could increase sense strand transcription and lead to the appearance of antisense down-regulation through transcriptional repression. This suggests that converging transcripts compete for transcriptional access to the same regions and that incorporation of 5' H2A.Z at a transcript's +1 nucleosome increases the likelihood of competitive transcription increasing from that locus and decreasing from the converging locus.

The S. cerevisiae study also examined antisense transcripts originating from the 3' ends of coding genes, but they looked specifically at the association between these transcripts and the H2A.Z levels in these same 3' locations. Here, H2A.Z had a positive affect on transcription originating from the locus. Both studies, therefore, are compatible with H2A.Z having a positive affect on whether or not transcriptional initiation occurs at a precise location. It should be noted, that, the median distance between *S. pombe* genes has been calculated as 442bp and while the distance between S. cerevisiae genes is 366bp (Zofall et al., 2009). As transcription produces different covalent modifications at different distances from the TSS (see Chapter 1), this variation in intergenic distances could also lead to different outcomes being produced from converging transcription.

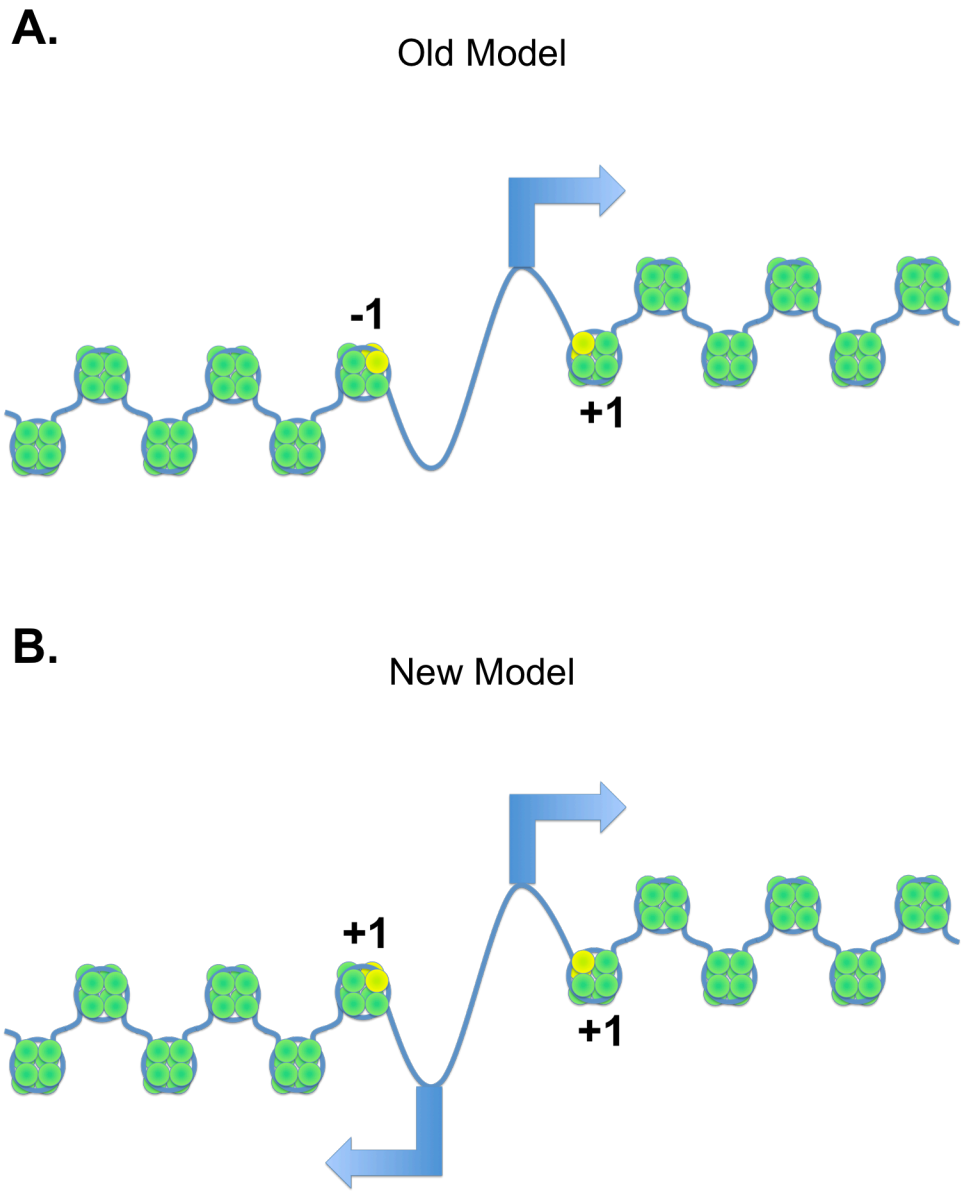

Figure 3.18: Models of H2A.Z Incorporation at the NFR

(A) The old model of H2A.Z incorporation at both the +1 and -1 nucleosomes at the NFR of a transcription start site.
(B) The new model demonstrating that incorporation of H2A.Z at both sides of an NFR is indicative of transcription produced in both directions. Hence, -1 nucleosomes that incorporate H2A.Z likely reflect +1 nucleosomes of diverging transcripts.

Chapter 4: H2A.Z and the Environment

"Summer surprised us, coming over the Starnbergersee"

-T. S. Elliot, *The Waste Land*

INTRODUCTION

One way to induce the remodeling of chromatin within yeast cultures is to subject cells to physiological stress. In particular, heat shock induced changes in gene expression have been shown to involve a repositioning of nucleosomes. In a 2008 paper, Shivaswamy et al. identified patterns of nucleosome displacement localized at gene promoters in response to heat shock (Shivaswamy et al., 2008). They found that nucleosome eviction was generally associated with gene activation while nucleosome appearance was generally associated with gene repression. This pattern makes sense, as strongly positioned nucleosomes could interfere with accessibility of DNA to the transcriptional machinery. There is, therefore, reason to think that the remodeling of local chromatin landscapes may play an integral part in cell responses to stress, ultimately contributing to the activation of groups of genes that respond to the specific stress conditions.

The deletion of some chromatin remodelers can also lead to changes in the location of nucleosomes. The deletion of the remodeler Chd1 has an impressive effect on global nucleosome positioning despite weak effects on phenotype (Gkikopoulos et al., 2011, Park et al., 2014b). However, in most cases single gene deletions do not show a dramatic effect on nucleosome positions. For example, deletion of the remodeler Snf2 has a dramatic phenotypic effect, significantly retarding cell growth (Shivaswamy and Iyer, 2008). However, when looking at the average transcription start site (TSS)

nucleosome profile, no large scale effects can be seen (data not shown). Since nucleosome positioning can be altered in a dynamic fashion by stimuli affecting large-scale gene expression, we wanted to investigate how H2A.Z and its chromatin remodelers might be involved in mediating appropriate chromatin responses to stress.

MATERIALS AND METHODS

Strains utilized

Cells used for the following experiments were all from the haploid BY4741 background (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0). Specific strains were obtained from various collections. A TAP-tagged H2A.Z strain was used from the TAP-tagged protein collection (Ghaemmaghami et al., 2003). Deletion mutant strains for Htz1 and Swr1 were obtained from the yeast haploid deletion strain collection (Open Biosystems, now GE Dharmacon) (Winzeler et al., 1999). Since the collection lacked an Ino80 deletion strain, we created the strain via tetrad dissection using heterozygous diploid cells containing the intact Ino80 gene and the deletion (Figure 4.1). We obtained this strain from the diploid essential deletion collection (from GE Dharmacon).

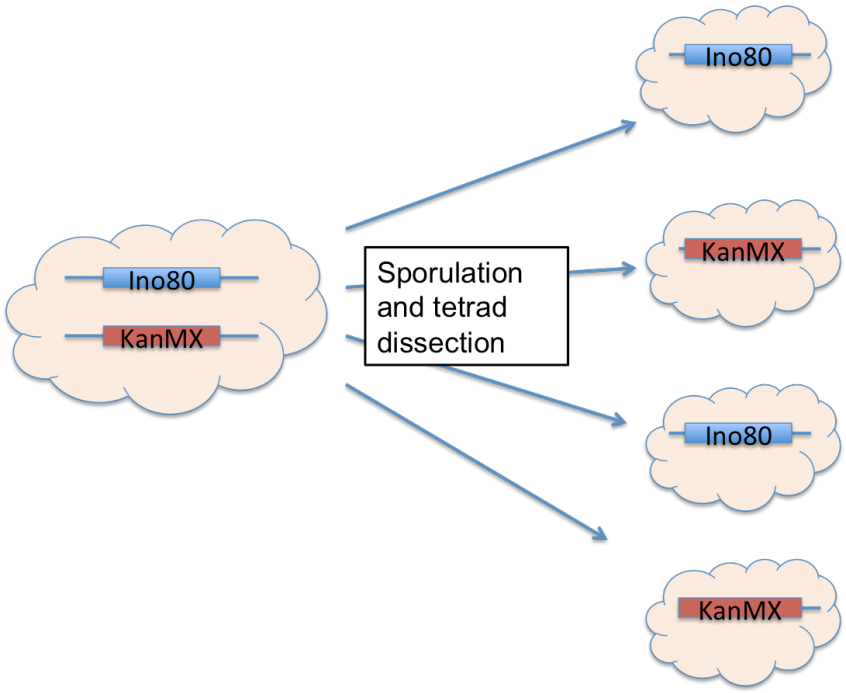

Figure 4.1: Schematic of haploid deletion strain creation by tetrad dissection

Diploid cells, heterozygous for the intact Ino80 gene and a deletion at that locus, were obtained from the diploid essential deletion collection (from GE Dharmacon), which is in the BY4743 background. The cells were transformed with a covering plasmid containing a copy of Ino80 to limit problems in meiosis. The diploid strains were then sporulated to produce tetrads. Tetrads were then dissected and only ones yielding 4 colonies were further screened. By replica plating on media containing Kan-MX, cells that possessed the KanMX cassette in place of the genomic copy of Ino80 were isolated. The cells were then grown without selection to give them a chance to lose the covering plasmid. Individual colonies were screened to select for plasmid loss. The resulting Ino80 haploid deletion strain was then confirmed with confirmation PCR.

Experimental conditions

Yeast cells were grown up in 200 ml cultures to an OD of .8. Cells were then divided into 4 (50ml) aliquots, and spun down at 30°C for 5mins at 4000rpm. The supernatants were discarded and cells were re-suspended in pre-warmed media, either at 30°C for normal (T0) or rapamycin treated samples, or at 39°C for heat-shocked samples.

The heat-shocked samples were then placed back into a 39°C shaking incubator. Rapamycin was added to the rapamycin treatment sample to a final concentration of 100nM, and the sample was placed in a 30°C shaking incubator for 30 minutes. The T0 sample was immediately fixed with formaldehyde. After 15 minute and 30 minute 39°C incubations, respectively, T15 and T30 heatshock samples were fixed and collected. All samples were collected after formaldehyde treatment by spinning down the samples, discarding the supernatant, and freezing the cell pellet for use downstream in MNase-seq or MNase ChIP-seq experiments (Figure 4.2). Since the 15 minute heatshock produced more pronounced changes in chromatin structure, we used this sample instead of the 30 minute heatshock sample for further downstream analysis.

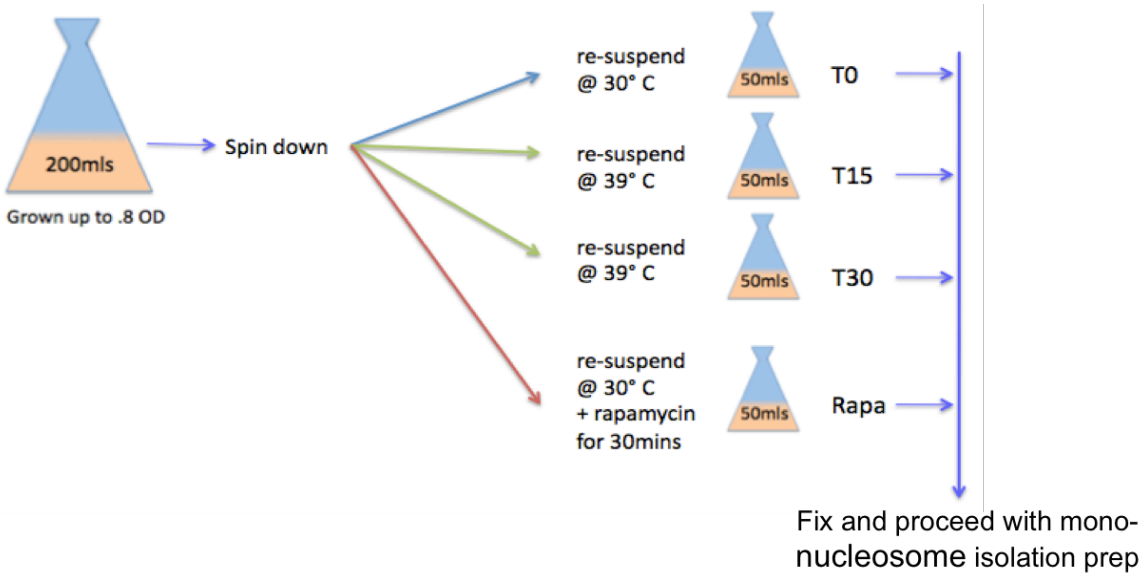

Figure 4.2: Schematic showing the experimental setup for heatshock and rapamycin treatment samples

MNase-seq and MNase ChIP-seq

The protocols used were identical to those described in Chapter 3.

Gene lists used for sorting

Lists of heat activated and heat repressed genes were obtained from Shivaswamy et al. which utilized microarray data from heat shocked cells to call differentially expressed genes with a log fold change cutoff of 2 and an adjusted p-value of .05 (Shivaswamy et al., 2008). Lists of rapamycin activated and rapamycin repressed genes were obtained from publicly available microarray data also using a log fold change cutoff of 2 and an adjusted p-value of .05 (Park et al., 2013).

RESULTS

We decided to characterize the effects of loss of H2A.Z and its chromatin remodelers on both phenotype and on chromatin structure. Deletion strains for Htz1 and for Swr1 were readily obtainable from the haploid deletion collection. However, since the Ino80 deletion mutant did not feature in the collection, we were forced to knock out the gene ourselves. Repeated failures in attempting to produce the knockout through conventional transformation methods, forced us to produce the strain using tetrad dissection (see Methods).

We first wanted to assess the affects of these deletions on growth rates of cells grown under typical laboratory conditions. This was performed both on solid media (with incubation in a standing 30°C incubator for 2 days) and in liquid culture (cells were grown in a shaking 30°C incubator with time-points tested for OD) (Figure 4.3). The ΔSwr1 strain showed the least amount of growth inhibition. Both ΔHtz1 and ΔIno80 showed pronounced growth defects, with ΔIno80's being the most severe. In light of the

difficulty in producing the ΔIno80 strain, the growth defect is not terribly surprising. The growth curve of the ΔIno80 strain suggests that, in addition to a longer doubling time, the mutant takes significantly longer to transition from stationary phase to log phase.

We next wanted to assess the distribution of nucleosomes within our deletion strains under normal growth conditions. Overall, nucleosomes around the NDR in the ΔHtz1 strain are still well positioned, indicating that H2A.Z is not required for producing regularly spaced arrays of nucleosomes within gene bodies (Figure 4.4). It has, previously, been argued that H2A.Z may affect positioning (Guillemette et al., 2005, Thakar et al., 2009) and that it does not (Li et al., 2005). This also indicates that decreased levels of H2A.Z at the NDR are not responsible for the fuzzy positioning of nucleosomes seen at genes with TATA containing promoters (Shivaswamy et al., 2008). By contrast, ΔIno80 displayed a dramatically altered nucleosome pattern. While the regular spacing between nucleosomes was well maintained, occupancy bordering the NDR and even within the NDR substantially increased. There was also a marked decrease in nucleosome occupancy levels through the gene body.

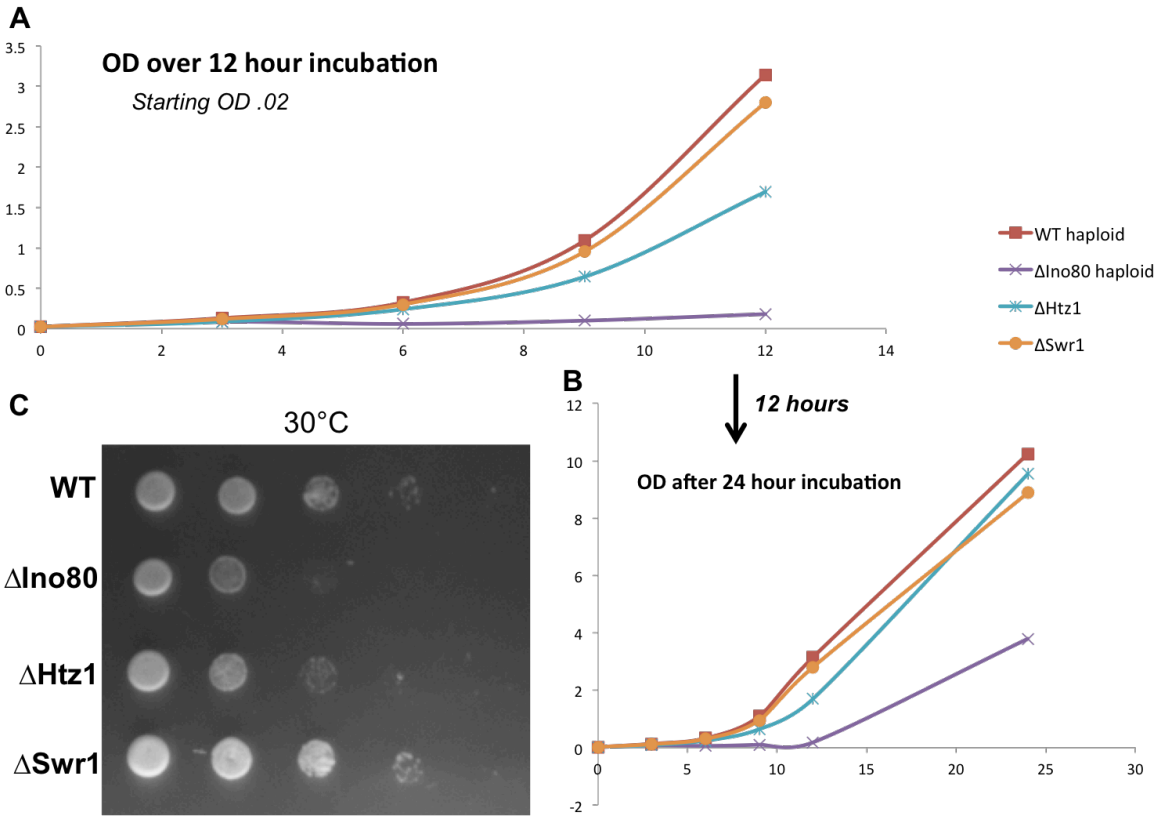

Figure 4.3: Growth defects of H2A.Z, Swr1, and Ino80 haploid deletion strains

(A) Growth curves, as measured by change in OD, of deletion strains grown in YPD with time-points taken every 3 hours. (B) Change in OD measured after 24 hours. (C) Growth defects of deletion strains on stationary media, assessed after 24 hours growth at 30°C. Serial dilutions were 1:10.

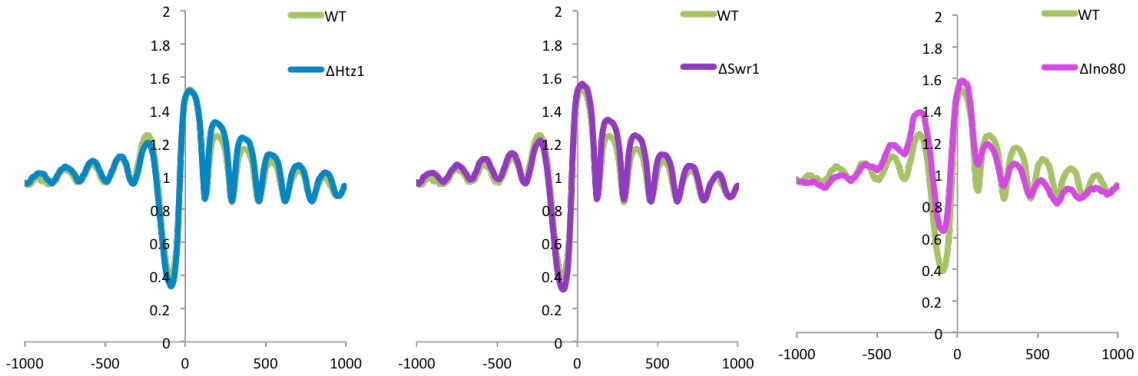

Figure 4.4: Average nucleosomes profiles across deletion strains

Mononucleosome profiles for each deletion mutant are plotted against the WT profiles. Overall, the Htz1 and Swr1 mutants produce little impact on global nucleosome patterns. By contrast, the Ino80 mutant exhibits reduced nucleosome levels throughout transcript bodies and increased levels in the NDR and -1 nucleosome position.

Heatshock

Evidence for the importance of H2A.Z in thermo-regulatory programs stems from experiments in several species. In yeast, deletion of Htz1 causes a small growth defect under normal conditions, but a much more pronounced one under heatshock (Santisteban et al., 2000). Additionally, H2A.Z incorporation has been found to increase the thermal mobility of nucleosomes, allowing for relocation at a temperature 4°C lower than nucleosomes containing canonical histones (Flaus et al., 2004). In Arabidopsis, under normal conditions, H2A.Z containing nucleosomes bind DNA tighter than canonical ones, but increasing temperatures lead to decreased H2A.Z nucleosome occupancy (Kumar and Wigge, 2010). Loss of H2A.Z causes these plants to grow as if they are grown at warm temperatures and to engage the transcription programs characteristic of those temperatures as well. Finally, in the fish *Cyprinus carpio,* H2A.Z levels within cells fluctuate with season (Talbert and Henikoff, 2014).

We wanted to further explore the effects of H2A.Z on heatshock tolerance, so, we subjected our deletion strains to a heatshock assay at 39°C (Figure 4.5). We found that heatshock inhibited growth for both the ΔHtz1 and the ΔIno80 strains, while having a comparatively small effect on ΔSwr1 cells. This result underscores the importance of H2A.Z in responding to a changing environment. The fact that the phenotype of the ΔSwr1 strain was so markedly different from the ΔHtz1 strain presents a few interesting possibilities. The first is that the rapid incorporation of H2A.Z at loci that are repressed can be accomplished in an Swr1 independent manner. The second is that the heat sensitivity is a consequence of problems caused by Swr1 and not by a lack of H2A.Z. Indeed, there is evidence, that, in the context of some other types of stresses, Swr1 causes genomic instability in the absence of H2A.Z, and that by deleting both the chromatin remodeler and histone variant these sensitivities can be ameliorated (Morillo-Huesca et al., 2010). The third, and more improbable, possibility is that H2A.Z performs a function related to heatshock response that is independent from its ability to be incorporated into chromatin. The extreme growth defect of the ΔIno80 strain suggests that an inability to remove H2A.Z from the genome is more problematic than an inability to incorporate H2A.Z

We next explored H2A.Z occupancy dynamics in WT cells at heat responsive loci. This was done by performing MNase ChIP for H2A.Z under normal conditions, and after a 15 minute heatshock at 39°C (Figure 4.6). We chose to focus our analyses on gene loci that are either up or down-regulated in response to heatshock in order to increase the likelihood of seeing occupancy changes that are transcriptionally relevant. In genes activated by heatshock, H2A.Z was depleted from all regions of our TSS profile (including upstream, -1, NDR, +1, and within the gene body). This finding is consistent with previous reports that H2A.Z is evicted during gene activation. For genes repressed

during heatshock, H2A.Z occupancy increased at the +1 nucleosome and within the gene body.

We then proceeded to characterize the effect of heatshock on the nucleosome profiles of our mutant strains at heat responsive genes. Overall, the profiles of our ΔSwr1 and ΔHtz1 strains did not different dramatically from WT. However, the ΔIno80 strain produced dramatically different results. Heat activated genes in this strain exhibited a severely dis-regulated nucleosome profile that largely seemed to have lost the regularly spaced nucleosomal arrays (Figure 4.7). These genes exhibited a dis-regulated pattern regardless of condition, but the shapes of the profiles between conditions did differ. Conversely, heat repressed genes in the Ino80 mutant exhibited more dis-regulated patterns under normal conditions and less dis-regulated patterns under heatshock (Figure 4.8). These results argue that the severe growth defect of the Ino80 mutant under heatshock conditions results from an inability to properly maintain nucleosome positioning and occupancy levels in the context of active transcription.

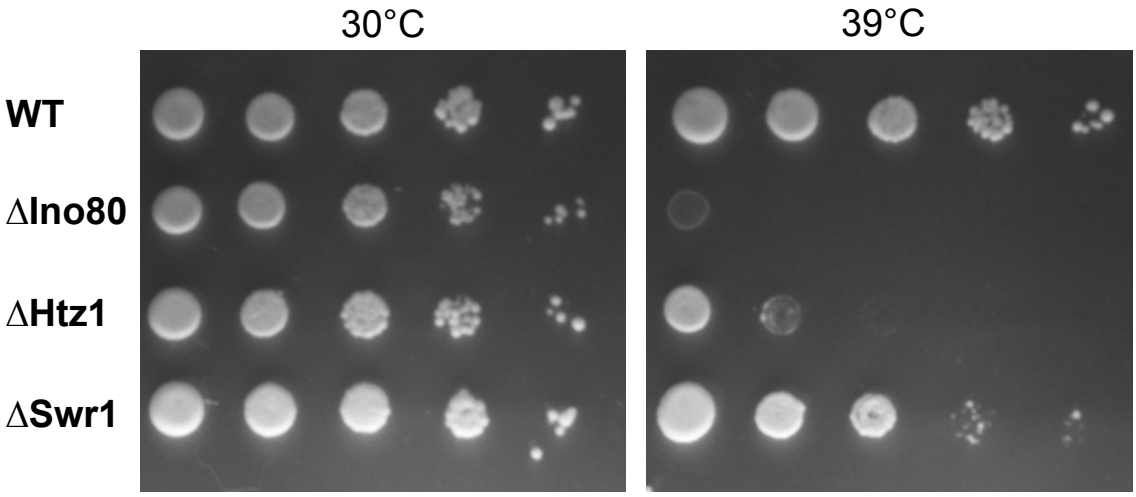

Figure 4.5: Phenotypic effects of deletion mutants exposed to heatshock

Strains were grown up overnight till saturation. They were then diluted to equivalent ODs and used for spotting assays, with serial dilutions of 1:10. Two identical plates were produced, one was placed in a 30°C incubator while the other was placed in a 39°C one. After two day's growth the 30°C plate was assessed for growth differences. Since growth at 39°C is considerably slower even in WT strains, the 39°C plate was assessed after 4 days.

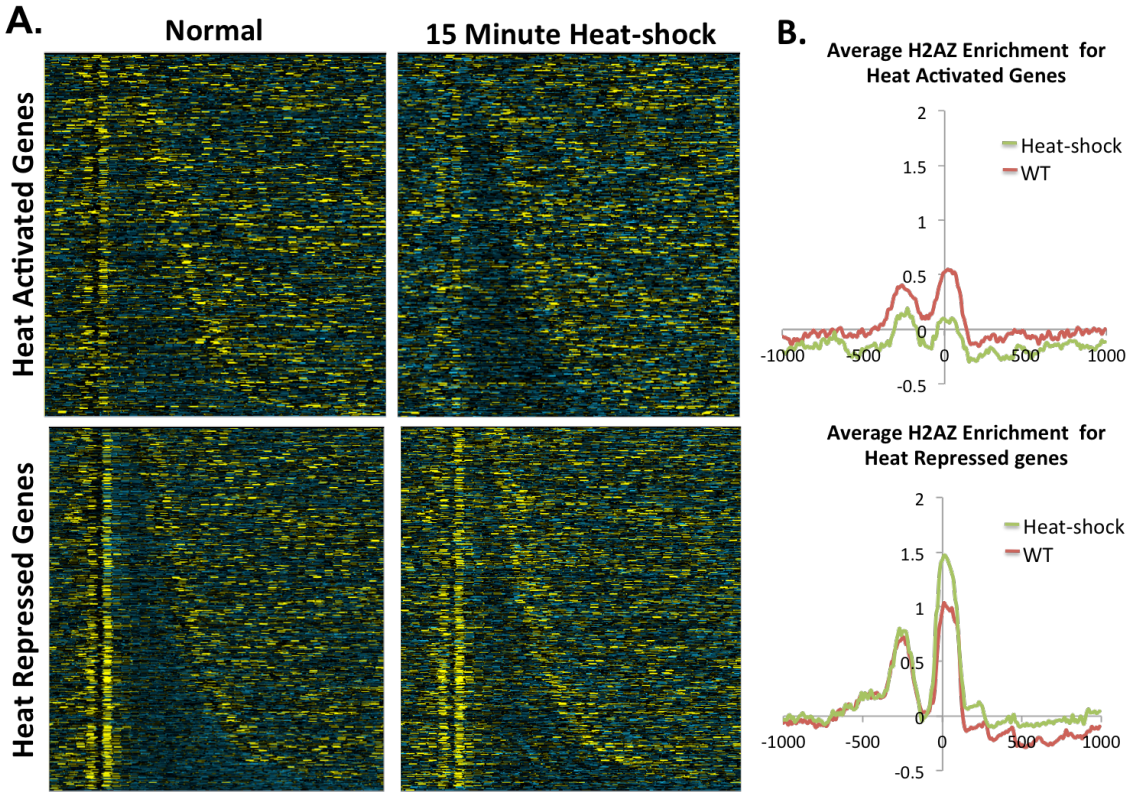

Figure 4.6: H2A.Z enrichment across heat responsive genes during heatshock

(A) Heatmaps produced with JavaTree displaying changes in H2A.Z localization between normal conditions and a 15 minute heatshock for heat activated and heat repressed genes. Genes are aligned by transcript length and display a window from -1000 bps upstream of the TSS to +6000 bps downstream. (B) Average profiles of H2A.Z enrichment across heat activated and heat repressed genes. Overall, actively transcribed genes are depleted for H2A.Z in both directions from their NDRs, whereas repressed genes accumulate H2A.Z only at the +1 nucleosome and within their gene bodies.

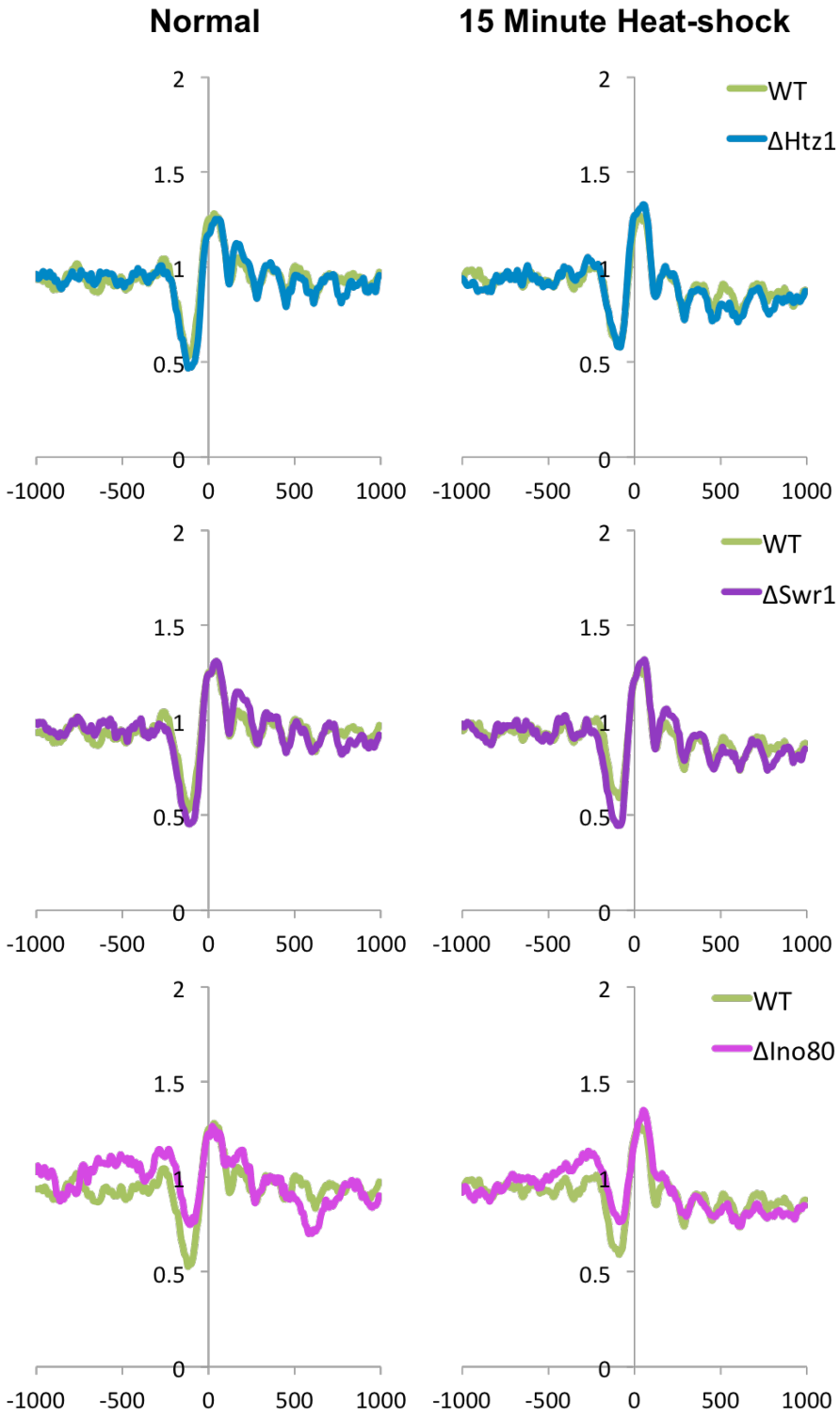

Figure 4.7: Average nucleosomes profiles across heat activated genes

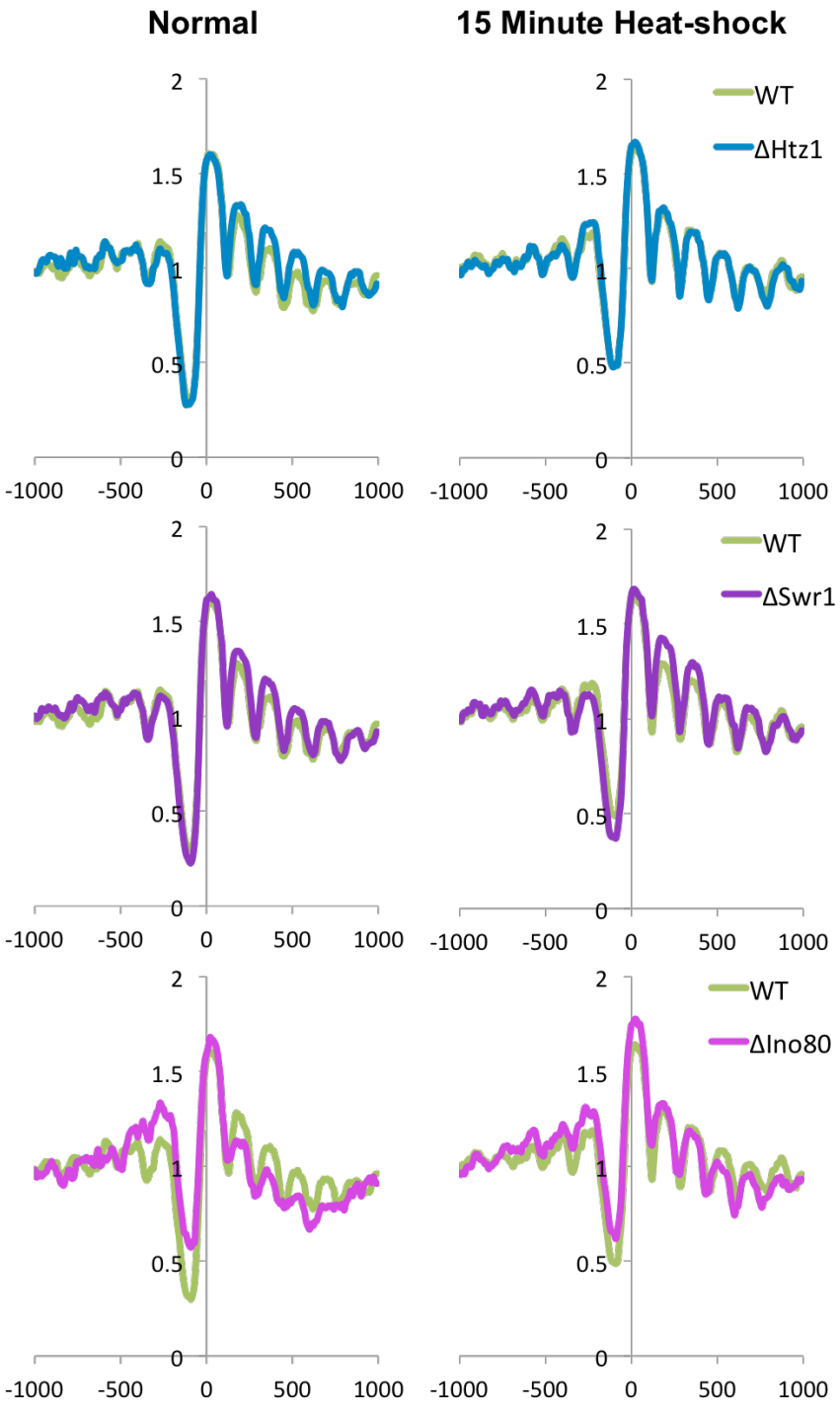

Figure 4.8: Average nucleosomes profiles across heat repressed genes

Rapamycin

We next turned our attention to characterizing H2A.Z based chromatin responses to the stress caused by rapamycin treatment. Phenotypic data suggests that rapamycin treatment induces G0 and mimics stationary phase (Zaragoza et al., 1998). Stationary phase cells display increased thermo-tolerance, maintaining viability at high temperatures for longer (Allen et al., 2006). In WT cells subjected to a 30 minute rapamycin treatment, we found that H2A.Z was depleted from the +1 nucleosome and from the gene body of activated genes (Figure 4.9). While this response was similar to the depletion seen in heat-shocked cells, the extent of the depletion was more modest, and did not include regions upstream of the +1 nucleosome. Rapamycin repressed genes accumulated H2A.Z within the gene body and slightly at the -1 nucleosome, but there was no increased incorporation at the +1 nucleosome. An explanation for the differences in the regions of H2A.Z incorporation and depletion between rapamycin responsive and heat responsive genes remains undetermined. These differences may reflect the activation of different chromatin regulatory modes, or perhaps result from the large scale repressive effects of rapamycin on gene transcription.

Turning our attention to the effects of rapamycin on nucleosome occupancy in our deletion strains, we observed that the ΔIno80 strain again displayed greater occupancy at the NDR and depletion distally, while the other two strains showed minimal differences compared to WT (Figure 4.10). Sorting the profiles into activated and repressed groups revealed similar global nucleosome patterns to those seen in our heat-shocked samples (Figure 4.11). Genes activated in response to rapamycin, displayed lower occupancy levels both before and after treatment than were seen for the genome at large, while repressed genes showed more nucleosome occupancy in both conditions. Ino80 patterns

were again aberrant, and most notably so for the activated genes both before and after treatment.

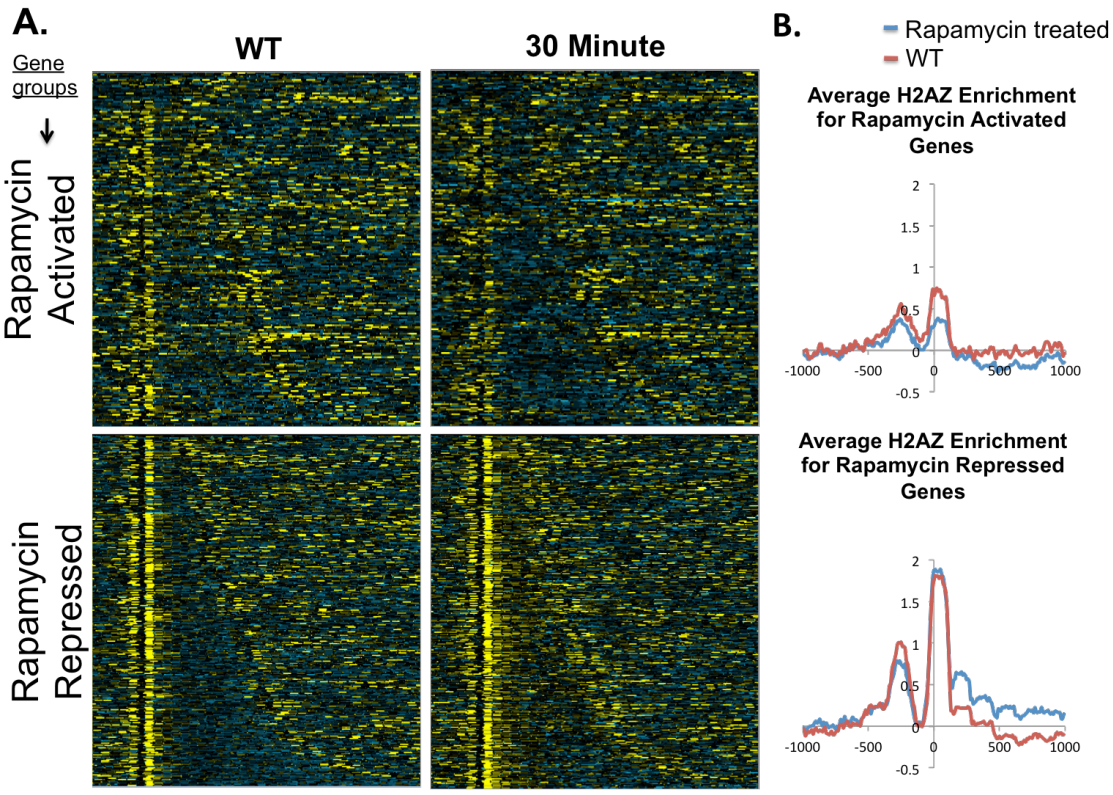

Figure 4.9: H2A.Z localization at rapamycin responsive genes under rapamycin treatment

(A) Heatmaps display changes in H2A.Z localization between normal conditions and a 30 minute rapamycin treatment for activated and repressed genes. Genes are aligned by transcript length and display a window from -1000 bps upstream of the TSS to +6000 bps downstream. (B) Average profiles of H2A.Z enrichment across rapamycin activated and repressed genes. Actively transcribed genes show H2A.Z eviction at the +1 nucleosome and near their 3' ends. Repressed genes incorporate H2A.Z within their gene bodies in regularly spaced arrays.

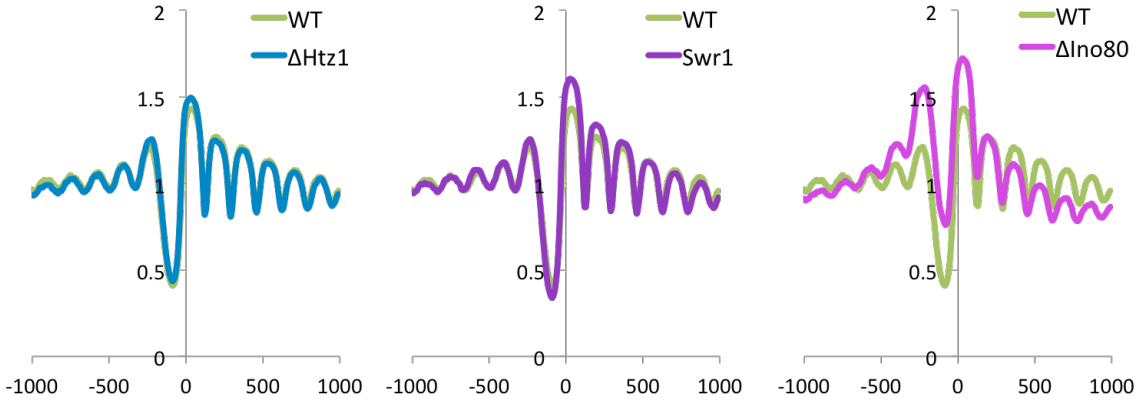

Figure 4.10: Average nucleosome positioning in rapamycin treated deletion strains across all genes

Overall, nucleosome patterns in the ΔHtz1 and ΔSwr1 strain are similar to WT patterns. By comparison, the ΔIno80 strain displays a very noticeably different pattern with nucleosomes over incorporated at the -1 and +1 nucleosome sites and mis-incorporated within the NDR.

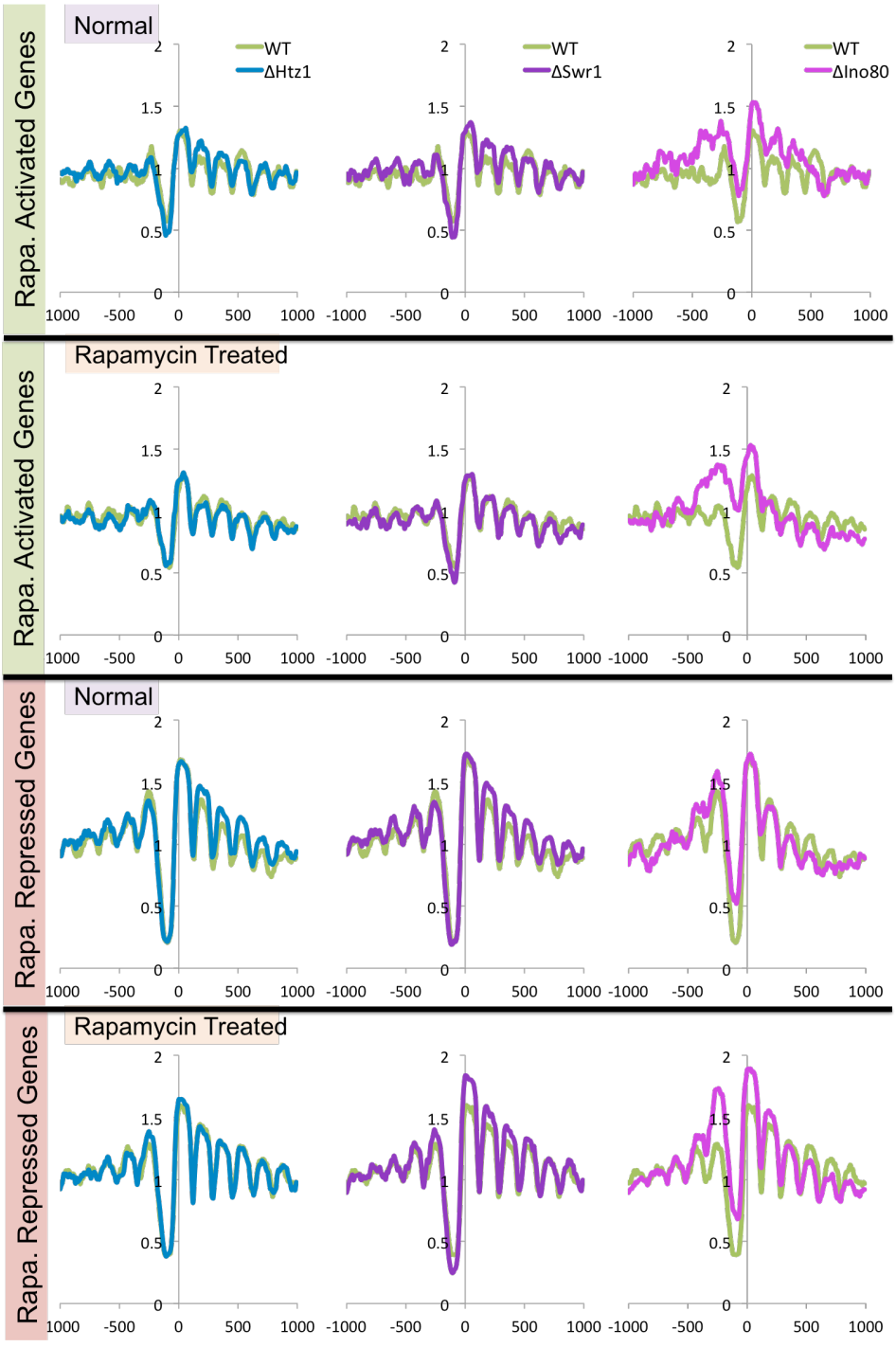

Figure 4.11: Nucleosome patterns across rapamycin responsive genes

Plotted above are average profiles of rapamycin activated and repressed genes across our samples. Rapamycin activated genes show aberrant nucleosome localization upstream of

(Figure 4.11 continued.) the TSS both in treated and non-treated samples within the ΔIno80 mutant strain. At rapamycin repressed genes, nucleosomes in the ΔIno80 mutant strain accumulate tightly at the +1 and -1 locations. WT cells instead showed decreased occupancies upstream of the -1 nucleosome and towards the 3' end for activated genes, and -1 depletions and 3' end increase for repressed genes.

RP genes

Finally, we decided to examine changes in nucleosome occupancy at the RP genes more closely, as, the expression patterns of these highly expressed genes have been shown to respond to cell stress. In particular, rapamycin induces down regulation of RP genes by inhibiting the target of rapamycin (TOR) pathway (Powers and Walter, 1999, Shamji et al., 2000). It was shown that for at least two of these RP genes, rapamycin caused the release of the histone acetyltransferase Esa1 from the locus, leading to decreased H4 acetylation levels (Rohde and Cardenas, 2003). Similarly, during heatshock, RP genes exhibit decreased H4 acetylation levels, and are generally repressed (Shivaswamy and Iyer, 2008). We therefore examined the nucleosome profiles of these genes in our data (Figure 4.12). Under both normal and rapamycin treatment conditions, deletion of Ino80 caused nucleosomes to increase in occupancy upstream of the RP genes. The size of the NDR in this strain was also noticeably compressed. By contrast, heatshock and rapamycin treatment of the ΔSwr1 strain caused an appreciable widening of, and decreased occupancy within, the NDR of RP genes. Deletion of Htz1 produced only minor changes in the nucleosome occupancy of these genes from WT.

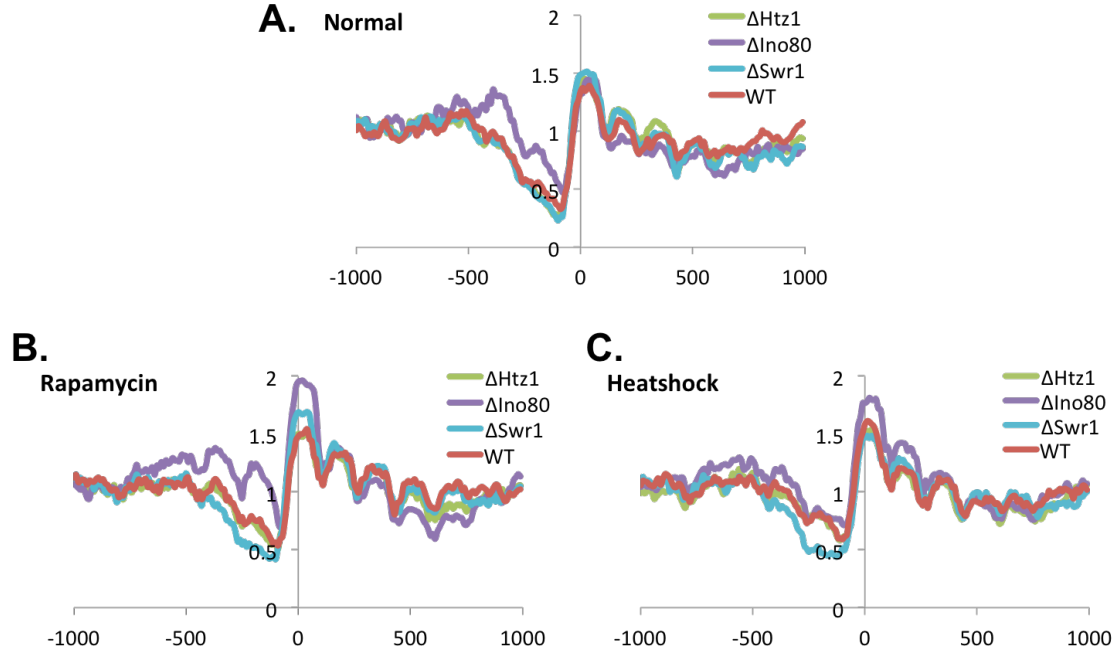

Figure 4.12: Nucleosome profiles across RP genes in deletion mutants and under stress

DISCUSSION

We have demonstrated that, in response to heatshock and rapamycin treatments, H2A.Z is depleted from activated genes and incorporated into repressed genes, though with somewhat different patterns. Previously, H2A.Z was shown to be incorporated more strongly into recently repressed genes (the study looked specifically at Ino1 and Gal1), and to be necessary for their rapid re-activation (Brickner et al., 2007). It may therefore serve the purpose of bookmarking a gene that was recently transcribed so that when conditions again require it, those genes may quickly be reactivated. This could provide a plausible reason why some genes lack H2A.Z at their NDRs, namely, that they are transcribed too infrequently to necessitate, or, perhaps, to allow for continued H2A.Z retention. In agreement with this model are studies showing that H2A.Z can act as a

binding platform for pioneer transcription factors, priming chromatin for future use (Subramanian et al., 2015). H2A.Z incorporation may, therefore, be a way to place a gene "on hold" so that it may then be retrieved faster at a later date. Evidence that H2A.Z plays a role in anti-silencing, and may prevent genes from being heterochromatinized (Kumar and Wigge, 2010) supports this theory.

We have also characterized the growth defects, thermo-tolerance sensitivities, and nucleosome profiles of strains deficient in aspects of H2A.Z regulation. In particular, we have shown that Ino80 deletion mutants take longer to transition from stationary phase to log phase, are incapable of handling heatshock, and produce highly aberrant nucleosome occupancy profiles. Nucleosomal dis-regulation is more pronounced at activated genes, and is present even before the administration of stress. A closer look reveals that genes activated in response to stress are highly enriched for TATA boxes, while genes that are repressed in response to stress are depleted for these promoter elements (Table 4.1). It therefore seems that, while H2A.Z is more strongly targeted to TATA-less promoters, its mis-regulation has a more detrimental impact on chromatin features at TATA containing genes where chromatin structure is less rigidly defined, and more susceptible to perturbation. It was shown in Chapter 3 that Ino80 can bind TATA box genes, and that its binding increases in the absence of H2A.Z. Our results, therefore, provide support for a model where H2A.Z is incorporated into TATA containing genes, and where Ino80 must actively remove it to maintain appropriate chromatin structure.

	Total Genes	TATA	TATA %	TATA-less	TATA-less %
Heat Activated	359	186	51.81	173	48.19
Heat Repressed	489	69	14.11	420	85.89
Rapamycin Activated	209	94	44.98	115	55.02
Rapamycin Repressed	302	20	6.62	282	93.38

Table 4.1: The proportion of stress responsive genes with TATA box containing promoters

CONCLUSION

H2A.Z has a long evolutionary history, likely having arisen only once, early in Eukaryotic evolution. It displays sequence conservation across species that is more strictly maintained than that for the canonical H2A histone (Thatcher and Gorovsky, 1994). Despite evidence of an ancient yeast genome duplication event (Wolfe and Shields, 1997), which could explain duplicate genes for canonical histones and their consistent pairings (Eriksson et al., 2012), in the yeast genome we find only one version of H2A.Z. These properties emphasis the important functions that this histone variant carries out, and how they must be strictly regulated and maintained.

Under "ideal" laboratory conditions, H2A.Z is dispensable for cell viability. In the wild, however, yeast cells must contend with a plethora of insults and challenges. Some of those most commonly experienced will, undoubtedly, be temperature fluctuations and nutrient depletion induced stationary phase growth. In these contexts, H2A.Z likely helps mediates the cells ability to rapidly remodel chromatin in order to activate frequently used transcriptional programs. By marking loci that have recently undergone transcription with H2A.Z, cells can, essentially, index these transcripts for future reference and prevent these loci from too quickly being sent back for heterochromatic storage. Conversely, dis-regulation in H2A.Z maintenance may lead to problems in responding appropriately to change. The disorganization of nucleosomal arrays seen within TATA genes in ΔIno80 cells, may reflect overuse or misuse of a cataloguing system meant instead to be used for functionally and structurally different TATA-less genes. Ultimately, along the yeast genome, NDRs are anchor points, setting the rhythm of well positioned proximal nucleosomes which fade with distance like

ripples over a pond. Nestled at their edges H2A.Z may be found providing signposts commemorating the occurrence and direction of recent transcriptional events.

Yeast provide us with a paired down nucleosome regulatory system, which, we can probe to gain insights about the basic functions of histones and chromatin remodelers. The complexity found within these components increases dramatically within metazoans. The ~88 human histone genes demonstrate how this complexity can increase. As the components of many of the human chromatin remodeling complexes have not yet been well defined, and may be highly tissue specific, a good deal of work remains to be done in order to translate insights gained from nucleosome regulation in yeast into insights in humans. However, investigating these proteins in humans may also provide some very different kinds of insights.

In order to more comprehensively study H2A.Z's functions within metazoans, and, more particularly, within humans, a number of techniques could be employed. Tissue specific immunoprecipitation followed by mass spectrometry could help clarify chromatin remodeler complex compositions. It may also help sort out interactions between the various human Swr1 and Ino80 homologs and the radiation of human H2A and H2A.Z histones. In addition, genome wide ChIP-seq experiments for these proteins could provide data on co-localized binding and more specific information about how these proteins actually interact within different tissue types.

Some components of the human H2A.Z regulatory system have already been revealed. The human Swr1 homologs that are responsible for H2A.Z deposition include the catalytic subunits SRCAP and EP400 (Gevry et al., 2009, Wong et al., 2007). The deposition chaperone YL1 may also contribute to H2A.Z incorporation (Latrick et al., 2016). Eviction of H2A.Z is likely accomplished by INO80 and the histone chaperone ANP32E (Obri et al., 2014).

Dis-regulation of chromatin is a common feature of cancers, and mutations in many histone variants and chromatin remodelers have been noted, including in H2A.Z variants and in its chromatin remodelers. Up-regulation of heat shock proteins has been noted in a number of cancers (Ciocca and Calderwood, 2005). In light of H2A.Z's conserved associations with heatshock response, a closer look at connections between changes in heat regulatory properties of cancer cells and possible mutations in these proteins is warranted. Of note, both hyperthermia and hypothermia have been used in cancer treatments with a number of positive effects in different cancers (Evans et al., 2015, Kalamida et al., 2015).

The Cancer Genome Atlas (TCGA) provides datasets detailing the numbers of somatic mutations found in specific genes across a variety of cancer types. Examining the mutation patterns within H2A.Z variants and some of its remodelers provides an interesting perspective (Figure C.1). Some cancers are strongly depleted for mutations in these genes, possibly suggesting that the tissue types they come from are highly dependent on these histones and complexes, and therefore extremely intolerant to mutations in them. What also becomes apparent, is that mutations in these genes show discernable patterns across cancer types. The commonalities in their mutation profiles, raises the possibility of using this type of mutation data to make predictions about gene and protein interactions. Ultimately, as more whole genome sequencing data becomes available, new approaches will allow us to uncover patterns and extract a wealth of information that could prove extremely useful in probing the molecular underpinnings of chromatin and transcription.

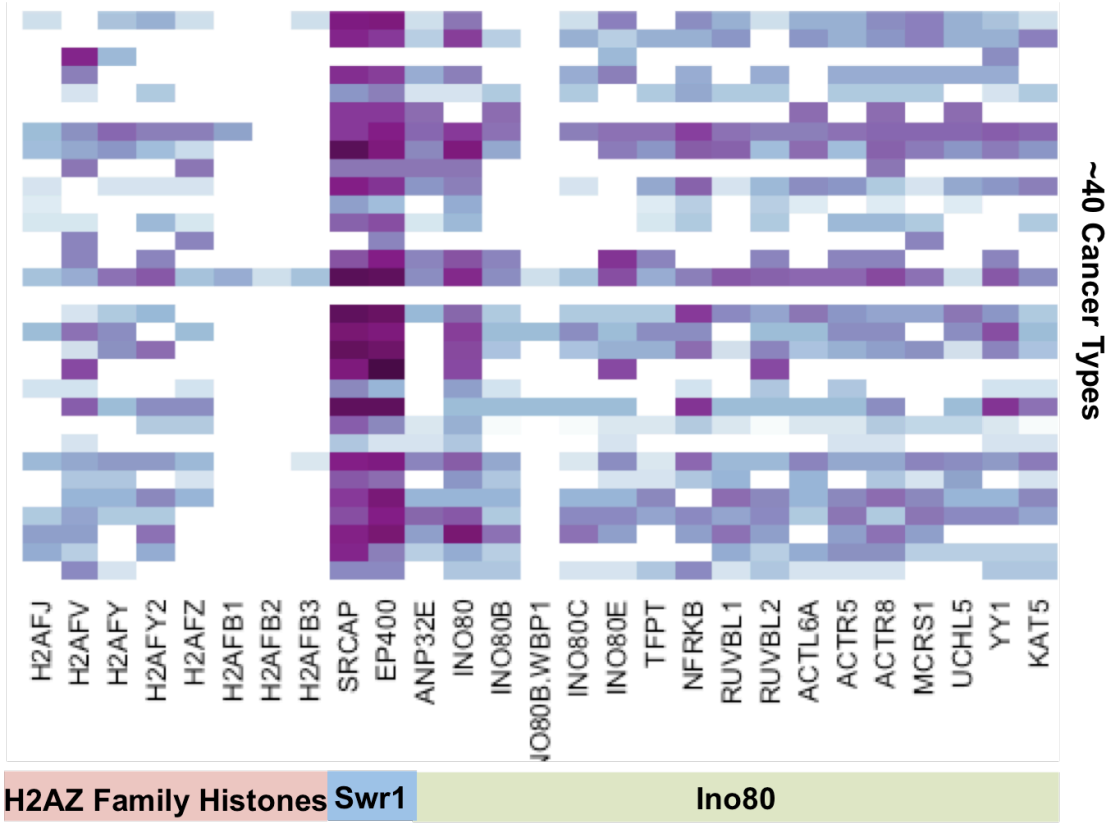

Figure C.1: Somatic mutations in H2A.Z and its chromatin remodelers in TCGA

Somatic mutations data was downloaded from TCGA in the form of .maf files. A master file was then created containing all calls of all types from all centers. Unique associations between patient IDs and genes with non-silent mutations were then derived. This provided us with a list of all genes with a mutation call for an individual patient. This list was then filtered by our list of H2A.Z associated genes within humans, and counts were separated by cancer type. All gene names listed to the right of Ino80 have been associated with the Ino80 complex. However, there is likely to be variation within these complexes between different tissue types, etc. RUVBL1 and RUVBL2 are also found within the SRCAP and EP400 complexes.

REFERENCES

1. Afek, A. and Lukatsky, D.B. (2013a) Genome-wide organization of eukaryotic preinitiation complex is influenced by nonconsensus protein-DNA binding. *Biophys J* 104, 1107-1115.
2. Afek, A. and Lukatsky, D.B. (2013b) Positive and negative design for nonconsensus protein-DNA binding affinity in the vicinity of functional binding sites. *Biophys J* 105, 1653-1660.
3. 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.
4. Albrecht, A.S. and Orom, U.A. (2015) Bidirectional expression of long ncRNA/protein-coding gene pairs in cancer. *Brief Funct Genomics*.
5. Allan, J., Hartman, P.G., Crane-Robinson, C., and Aviles, F.X. (1980) The structure of histone H1 and its location in chromatin. *Nature* 288, 675-679.
6. Allen, C., Buttner, S., Aragon, A.D., Thomas, J.A., Meirelles, O., Jaetao, J.E., . . . Werner-Washburne, M. (2006) Isolation of quiescent and nonquiescent cells from yeast stationary-phase cultures. *J Cell Biol* 174, 89-100.
7. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* 499, 360-363.
8. Anderson, S.K. (2014) Probabilistic bidirectional promoter switches: noncoding RNA takes control. *Mol Ther Nucleic Acids* 3, e191.
9. Andersson, R., Chen, Y., Core, L., Lis, J.T., Sandelin, A., and Jensen, T.H. (2015a) Human Gene Promoters Are Intrinsically Bidirectional. *Mol Cell* 60, 346-347.
10. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., . . . Sandelin, A. (2014) An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455-461.
11. Andersson, R., Sandelin, A., and Danko, C.G. (2015b) A unified architecture of transcriptional regulatory elements. *Trends Genet* 31, 426-433.
12. Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drablos, F., . . . Hayashizaki, Y. (2015) Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* 347, 1010-1014.
13. Audia, J.E. and Campbell, R.M. (2016) Histone Modifications and Cancer. *Cold Spring Harb Perspect Biol* 8.
14. Bagchi, D.N. and Iyer, V.R. (2016) The Determinants of Directionality in Transcriptional Initiation. *Trends Genet*.
15. Balbin, O.A., Malik, R., Dhanasekaran, S.M., Prensner, J.R., Cao, X., Wu, Y.M., . . . Chinnaiyan, A.M. (2015) The landscape of antisense gene expression in human cancers. *Genome Res* 25, 1068-1079.

16. Baltimore, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 226, 1209-1211.
17. Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., . . . Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823-837.
18. 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.
19. Birnbaum, R.Y., Clowney, E.J., Agamy, O., Kim, M.J., Zhao, J., Yamanaka, T., . . . Ahituv, N. (2012) Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res* 22, 1059-1068.
20. Blazie, S.M., Babb, C., Wilky, H., Rawls, A., Park, J.G., and Mangone, M. (2015) Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in Caenorhabditis elegans intestine and muscles. *BMC Biol* 13, 4.
21. Bloom, K. and Joglekar, A. (2010) Towards building a chromosome segregation machine. *Nature* 463, 446-456.
22. Bonisch, C. and Hake, S.B. (2012) Histone H2A variants in nucleosomes and chromatin: more or less stable? *Nucleic Acids Res* 40, 10719-10741.
23. Bornelov, S., Komorowski, J., and Wadelius, C. (2015) Different distribution of histone modifications in genes with unidirectional and bidirectional transcription and a role of CTCF and cohesin in directing transcription. *BMC Genomics* 16, 300.
24. Brickner, D.G., Cajigas, I., Fondufe-Mittendorf, Y., Ahmed, S., Lee, P.C., Widom, J., and Brickner, J.H. (2007) H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS Biol* 5, e81.
25. Brown, S.J., Stoilov, P., and Xing, Y. (2012) Chromatin and epigenetic regulation of pre-mRNA processing. *Hum Mol Genet* 21, R90-96.
26. Camblong, J., Beyrouthy, N., Guffanti, E., Schlaepfer, G., Steinmetz, L.M., and Stutz, F. (2009) Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. *Genes Dev* 23, 1534-1545.
27. Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., and Stutz, F. (2007) Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. *Cell* 131, 706-717.
28. Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., . . . Hayashizaki, Y. (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 38, 626-635.
29. Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., . . . Workman, J.L. (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123, 581-592.
30. Cassiday, L.A. and Maher, L.J., 3rd (2002) Having it both ways: transcription factors that bind DNA and RNA. *Nucleic Acids Res* 30, 4118-4126.

31. Castelnuovo, M., Zaugg, J.B., Guffanti, E., Maffioletti, A., Camblong, J., Xu, Z., . . . Stutz, F. (2014) Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. *Nucleic Acids Res* 42, 4348-4362.
32. Chang, D.T., Wu, C.Y., and Fan, C.Y. (2012) A study on promoter characteristics of head-to-head genes in Saccharomyces cerevisiae. *BMC Genomics* 13 Suppl 1, S11.
33. Chen, P., Zhao, J., Wang, Y., Wang, M., Long, H., Liang, D., . . . Li, G. (2013) H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. *Genes Dev* 27, 2109-2124.
34. Churchman, L.S. and Weissman, J.S. (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368-373.
35. Ciocca, D.R. and Calderwood, S.K. (2005) Heat shock proteins in cancer: diagnostic, prognostic, predictive, and treatment implications. *Cell Stress Chaperones* 10, 86-103.
36. Collins, P.J., Kobayashi, Y., Nguyen, L., Trinklein, N.D., and Myers, R.M. (2007) The ets-related transcription factor GABP directs bidirectional transcription. *PLoS Genet* 3, e208.
37. Consortium, E.P., Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74.
38. Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014) Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* 46, 1311-1320.
39. Core, L.J., Waterfall, J.J., Gilchrist, D.A., Fargo, D.C., Kwak, H., Adelman, K., and Lis, J.T. (2012) Defining the status of RNA polymerase at promoters. *Cell Rep* 2, 1025-1035.
40. Core, L.J., Waterfall, J.J., and Lis, J.T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845-1848.
41. Crick, F. (1970) Central dogma of molecular biology. *Nature* 227, 561-563.
42. Danielsson, F., James, T., Gomez-Cabrero, D., and Huss, M. (2015) Assessing the consistency of public human tissue RNA-seq data sets. *Brief Bioinform* 16, 941-949.
43. de Hoon, M. and Hayashizaki, Y. (2008) Deep cap analysis gene expression (CAGE): genome-wide identification of promoters, quantification of their expression, and network inference. *Biotechniques* 44, 627-628, 630, 632.
44. 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.
45. Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., . . . Gingeras, T.R. (2012) Landscape of transcription in human cells. *Nature* 489, 101-108.

46. Draker, R., Sarcinella, E., and Cheung, P. (2011) USP10 deubiquitylates the histone variant H2A.Z and both are required for androgen receptor-mediated gene activation. *Nucleic Acids Res* 39, 3529-3542.
47. Drouin, J. (2014) Minireview: pioneer transcription factors in cell fate specification. *Mol Endocrinol* 28, 989-998.
48. Duttke, S.H., Lacadie, S.A., Ibrahim, M.M., Glass, C.K., Corcoran, D.L., Benner, C., . . . Ohler, U. (2015) Human promoters are intrinsically directional. *Mol Cell* 57, 674-684.
49. Engstrom, P.G., Ho Sui, S.J., Drivenes, O., Becker, T.S., and Lenhard, B. (2007) Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res* 17, 1898-1908.
50. Eriksson, P.R., Ganguli, D., Nagarajavel, V., and Clark, D.J. (2012) Regulation of histone gene expression in budding yeast. *Genetics* 191, 7-20.
51. Evans, S.S., Repasky, E.A., and Fisher, D.T. (2015) Fever and the thermal regulation of immunity: the immune system feels the heat. *Nat Rev Immunol* 15, 335-349.
52. Fang, R., Moss, W.N., Rutenberg-Schoenberg, M., and Simon, M.D. (2015) Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. *PLoS Genet* 11, e1005668.
53. FitzGerald, P.C., Sturgill, D., Shyakhtenko, A., Oliver, B., and Vinson, C. (2006) Comparative genomics of Drosophila and human core promoters. *Genome Biol* 7, R53.
54. Flaus, A., Rencurel, C., Ferreira, H., Wiechens, N., and Owen-Hughes, T. (2004) Sin mutations alter inherent nucleosome mobility. *EMBO J* 23, 343-353.
55. Gatta, R. and Mantovani, R. (2011) NF-Y affects histone acetylation and H2A.Z deposition in cell cycle promoters. *Epigenetics* 6, 526-534.
56. Gevry, N., Hardy, S., Jacques, P.E., Laflamme, L., Svotelis, A., Robert, F., and Gaudreau, L. (2009) Histone H2A.Z is essential for estrogen receptor signaling. *Genes Dev* 23, 1522-1533.
57. Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., . . . Weissman, J.S. (2003) Global analysis of protein expression in yeast. *Nature* 425, 737-741.
58. Gkikopoulos, T., Schofield, P., Singh, V., Pinskaya, M., Mellor, J., Smolle, M., . . . Owen-Hughes, T. (2011) A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. *Science* 333, 1758-1760.
59. Gray, K.A., Seal, R.L., Tweedie, S., Wright, M.W., and Bruford, E.A. (2016) A review of the new HGNC gene family resource. *Hum Genomics* 10, 6.
60. Gu, M., Naiyachit, Y., Wood, T.J., and Millar, C.B. (2015) H2A.Z marks antisense promoters and has positive effects on antisense transcript levels in budding yeast. *BMC Genomics* 16, 99.
61. Guillemette, B., Bataille, A.R., Gevry, N., Adam, M., Blanchette, M., Robert, F., and Gaudreau, L. (2005) Variant histone H2A.Z is globally localized to the

promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol* 3, e384.
62. Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., . . . Lenhard, B. (2014) Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature* 507, 381-385.
63. Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013) Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res* 23, 1210-1223.
64. Hardy, S., Jacques, P.E., Gevry, N., Forest, A., Fortin, M.E., Laflamme, L., . . . Robert, F. (2009) The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition. *PLoS Genet* 5, e1000687.
65. Hardy, S. and Robert, F. (2010) Random deposition of histone variants: A cellular mistake or a novel regulatory mechanism? *Epigenetics* 5, 368-372.
66. Hennig, B.P., Bendrin, K., Zhou, Y., and Fischer, T. (2012) Chd1 chromatin remodelers maintain nucleosome organization and repress cryptic transcription. *EMBO Rep* 13, 997-1003.
67. Heo, J.B. and Sung, S. (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 331, 76-79.
68. 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.
69. Hou, C., Zhao, H., Tanimoto, K., and Dean, A. (2008) CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc Natl Acad Sci U S A* 105, 20398-20403.
70. Hsieh, C.L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., . . . Kantoff, P.W. (2014) Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc Natl Acad Sci U S A* 111, 7319-7324.
71. Hu, G., Cui, K., Northrup, D., Liu, C., Wang, C., Tang, Q., . . . Zhao, K. (2013) H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* 12, 180-192.
72. Huang, W., Wong, J.M., and Bateman, E. (1996) TATA elements direct bi-directional transcription by RNA polymerases II and III. *Nucleic Acids Res* 24, 1158-1163.
73. Ietswaart, R., Wu, Z., and Dean, C. (2012) Flowering time control: another window to the connection between antisense RNA and chromatin. *Trends Genet* 28, 445-453.
74. Iouzalen, N., Moreau, J., and Mechali, M. (1996) H2A.ZI, a new variant histone expressed during Xenopus early development exhibits several distinct features from the core histone H2A. *Nucleic Acids Res* 24, 3947-3952.
75. Iwafuchi-Doi, M. and Zaret, K.S. (2014) Pioneer transcription factors in cell reprogramming. *Genes Dev* 28, 2679-2692.

76. Jiang, C. and Pugh, B.F. (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 10, 161-172.
77. Jin, C. and Felsenfeld, G. (2007) Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev* 21, 1519-1529.
78. Joseph, R., Orlov, Y.L., Huss, M., Sun, W., Kong, S.L., Ukil, L., . . . Liu, E.T. (2010) Integrative model of genomic factors for determining binding site selection by estrogen receptor-alpha. *Mol Syst Biol* 6, 456.
79. Kalamida, D., Karagounis, I.V., Mitrakas, A., Kalamida, S., Giatromanolaki, A., and Koukourakis, M.I. (2015) Fever-range hyperthermia vs. hypothermia effect on cancer cell viability, proliferation and HSP90 expression. *PLoS One* 10, e0116021.
80. Kamakaka, R.T. and Biggins, S. (2005) Histone variants: deviants? *Genes Dev* 19, 295-310.
81. Kaplan, C.D., Laprade, L., and Winston, F. (2003) Transcription elongation factors repress transcription initiation from cryptic sites. *Science* 301, 1096-1099.
82. Kaplan, T., Li, X.Y., Sabo, P.J., Thomas, S., Stamatoyannopoulos, J.A., Biggin, M.D., and Eisen, M.B. (2011) Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early Drosophila development. *PLoS Genet* 7, e1001290.
83. Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423, 241-254.
84. Keogh, M.C., Mennella, T.A., Sawa, C., Berthelet, S., Krogan, N.J., Wolek, A., . . . Buratowski, S. (2006) The Saccharomyces cerevisiae histone H2A variant Htz1 is acetylated by NuA4. *Genes Dev* 20, 660-665.
85. Kim, T. and Buratowski, S. (2009) Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell* 137, 259-272.
86. Kim, T., Xu, Z., Clauder-Munster, S., Steinmetz, L.M., and Buratowski, S. (2012) Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics. *Cell* 150, 1158-1169.
87. Kowalczyk, M.S., Hughes, J.R., Garrick, D., Lynch, M.D., Sharpe, J.A., Sloane-Stanley, J.A., . . . Higgs, D.R. (2012) Intragenic enhancers act as alternative promoters. *Mol Cell* 45, 447-458.
88. Kruesi, W.S., Core, L.J., Waters, C.T., Lis, J.T., and Meyer, B.J. (2013) Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *Elife* 2, e00808.
89. Kumar, S.V. and Wigge, P.A. (2010) H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. *Cell* 140, 136-147.
90. Kutach, A.K. and Kadonaga, J.T. (2000) The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Mol Cell Biol* 20, 4754-4764.

91. Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013) Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* 339, 950-953.
92. Lantermann, A.B., Straub, T., Stralfors, 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.
93. Latrick, C.M., Marek, M., Ouararhni, K., Papin, C., Stoll, I., Ignatyeva, M., . . . Hamiche, A. (2016) Molecular basis and specificity of H2A.Z-H2B recognition and deposition by the histone chaperone YL1. *Nat Struct Mol Biol* 23, 309-316.
94. Leung, D., Jung, I., Rajagopal, N., Schmitt, A., Selvaraj, S., Lee, A.Y., . . . Ren, B. (2015) Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* 518, 350-354.
95. Lewin, R. (1982) RNA can be a catalyst. *Science* 218, 872-874.
96. Li, B., Pattenden, S.G., Lee, D., Gutierrez, J., Chen, J., Seidel, C., . . . 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.
97. Li, W., Lam, M.T., and Notani, D. (2014) Enhancer RNAs. *Cell Cycle* 13, 3151-3152.
98. Lickwar, C.R., Rao, B., Shabalin, A.A., Nobel, A.B., Strahl, B.D., and Lieb, J.D. (2009) The Set2/Rpd3S pathway suppresses cryptic transcription without regard to gene length or transcription frequency. *PLoS One* 4, e4886.
99. Lin, J.M., Collins, P.J., Trinklein, N.D., Fu, Y., Xi, H., Myers, R.M., and Weng, Z. (2007) Transcription factor binding and modified histones in human bidirectional promoters. *Genome Res* 17, 818-827.
100. Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., . . . Pringle, J.R. (1998) Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* 14, 953-961.
101. Lubliner, S., Regev, I., Lotan-Pompan, M., Edelheit, S., Weinberger, A., and Segal, E. (2015) Core promoter sequence in yeast is a major determinant of expression level. *Genome Res* 25, 1008-1017.
102. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389, 251-260.
103. Marquardt, S., Escalante-Chong, R., Pho, N., Wang, J., Churchman, L.S., Springer, M., and Buratowski, S. (2014) A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell* 157, 1712-1723.
104. Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J., and Maltais, L.J. (2002) The human and mouse replication-dependent histone genes. *Genomics* 80, 487-498.

105. Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., . . . Pugh, B.F. (2008) Nucleosome organization in the Drosophila genome. *Nature* 453, 358-362.
106. Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., . . . Churchman, L.S. (2015) Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* 161, 541-554.
107. Maze, I., Noh, K.M., Soshnev, A.A., and Allis, C.D. (2014) Every amino acid matters: essential contributions of histone variants to mammalian development and disease. *Nat Rev Genet* 15, 259-271.
108. Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Oude Vrielink, J.A., . . . Agami, R. (2013) eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol Cell* 49, 524-535.
109. Meneghini, M.D., Wu, M., and Madhani, H.D. (2003) Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* 112, 725-736.
110. Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., . . . Bernstein, B.E. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553-560.
111. Millar, C.B., Xu, F., Zhang, K., and Grunstein, M. (2006) Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast. *Genes Dev* 20, 711-722.
112. Mizuguchi, G., Shen, X., Landry, J., Wu, W.H., Sen, S., and Wu, C. (2004) ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* 303, 343-348.
113. Morillo-Huesca, M., Clemente-Ruiz, M., Andujar, E., and Prado, F. (2010) The SWR1 histone replacement complex causes genetic instability and genome-wide transcription misregulation in the absence of H2A.Z. *PLoS One* 5, e12143.
114. Nagawa, F. and Fink, G.R. (1985) The relationship between the "TATA" sequence and transcription initiation sites at the HIS4 gene of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 82, 8557-8561.
115. Nakao, A., Yoshihama, M., and Kenmochi, N. (2004) RPG: the Ribosomal Protein Gene database. *Nucleic Acids Res* 32, D168-170.
116. Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457, 1038-1042.
117. Nelson, H.C., Finch, J.T., Luisi, B.F., and Klug, A. (1987) The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* 330, 221-226.
118. Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., . . . Shendure, J. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272-276.
119. Nojima, T., Gomes, T., Grosso, A.R., Kimura, H., Dye, M.J., Dhir, S., . . . Proudfoot, N.J. (2015) Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* 161, 526-540.

120. Noller, H.F., Hoffarth, V., and Zimniak, L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* 256, 1416-1419.
121. Ntini, E., Jarvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jorgensen, M., . . . Jensen, T.H. (2013) Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol* 20, 923-928.
122. Obri, A., Ouararhni, K., Papin, C., Diebold, M.L., Padmanabhan, K., Marek, M., . . . Hamiche, A. (2014) ANP32E is a histone chaperone that removes H2A.Z from chromatin. *Nature* 505, 648-653.
123. Olins, A.L. and Olins, D.E. (1974) Spheroid chromatin units (v bodies). *Science* 183, 330-332.
124. Papamichos-Chronakis, M., Watanabe, S., Rando, O.J., and Peterson, C.L. (2011) Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* 144, 200-213.
125. Park, D., Lee, Y., Bhupindersingh, G., and Iyer, V.R. (2013) Widespread misinterpretable ChIP-seq bias in yeast. *PLoS One* 8, e83506.
126. Park, D., Morris, A.R., Battenhouse, A., and Iyer, V.R. (2014a) Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. *Nucleic Acids Res* 42, 3736-3749.
127. Park, D., Shivram, H., and Iyer, V.R. (2014b) Chd1 co-localizes with early transcription elongation factors independently of H3K36 methylation and releases stalled RNA polymerase II at introns. *Epigenetics Chromatin* 7, 32.
128. Pelechano, V., Wei, W., and Steinmetz, L.M. (2013) Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* 497, 127-131.
129. Powers, T. and Walter, P. (1999) Regulation of ribosome biogenesis by the rapamycin-sensitive TOR-signaling pathway in Saccharomyces cerevisiae. *Mol Biol Cell* 10, 987-1000.
130. Preker, P., Almvig, K., Christensen, M.S., Valen, E., Mapendano, C.K., Sandelin, A., and Jensen, T.H. (2011) PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res* 39, 7179-7193.
131. Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., . . . Jensen, T.H. (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322, 1851-1854.
132. Rhee, H.S. and Pugh, B.F. (2012) Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* 483, 295-301.
133. Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., . . . Chang, H.Y. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311-1323.
134. Ritter, D.I., Dong, Z., Guo, S., and Chuang, J.H. (2012) Transcriptional enhancers in protein-coding exons of vertebrate developmental genes. *PLoS One* 7, e35202.

135. Rohde, J.R. and Cardenas, M.E. (2003) The tor pathway regulates gene expression by linking nutrient sensing to histone acetylation. *Mol Cell Biol* 23, 629-635.
136. Santisteban, M.S., Hang, M., and Smith, M.M. (2011) Histone variant H2A.Z and RNA polymerase II transcription elongation. *Mol Cell Biol* 31, 1848-1860.
137. Santisteban, M.S., Kalashnikova, T., and Smith, M.M. (2000) Histone H2A.Z regulats transcription and is partially redundant with nucleosome remodeling complexes. *Cell* 103, 411-422.
138. Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012) The long-range interaction landscape of gene promoters. *Nature* 489, 109-113.
139. Sarcinella, E., Zuzarte, P.C., Lau, P.N., Draker, R., and Cheung, P. (2007) Monoubiquitylation of H2A.Z distinguishes its association with euchromatin or facultative heterochromatin. *Mol Cell Biol* 27, 6457-6468.
140. Scruggs, B.S., Gilchrist, D.A., Nechaev, S., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2015) Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Mol Cell* 58, 1101-1112.
141. Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., . . . Sharp, P.A. (2008) Divergent transcription from active promoters. *Science* 322, 1849-1851.
142. Shamji, A.F., Kuruvilla, F.G., and Schreiber, S.L. (2000) Partitioning the transcriptional program induced by rapamycin among the effectors of the Tor proteins. *Curr Biol* 10, 1574-1581.
143. Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., . . . Segal, E. (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat Biotechnol* 30, 521-530.
144. Sherwood, R.I., Hashimoto, T., O'Donnell, C.W., Lewis, S., Barkal, A.A., van Hoff, J.P., . . . Gifford, D.K. (2014) Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat Biotechnol* 32, 171-178.
145. Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V.R. (2008) Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol* 6, e65.
146. Shivaswamy, S. and Iyer, V.R. (2008) Stress-dependent dynamics of global chromatin remodeling in yeast: dual role for SWI/SNF in the heat shock stress response. *Mol Cell Biol* 28, 2221-2234.
147. Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., . . . Young, R.A. (2015) Transcription factor trapping by RNA in gene regulatory elements. *Science* 350, 978-981.
148. Skene, P.J., Hernandez, A.E., Groudine, M., and Henikoff, S. (2014) The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. *Elife* 3, e02042.

149. Snyder, M.W., Kircher, M., Hill, A.J., Daza, R.M., and Shendure, J. (2016) Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* 164, 57-68.
150. Soufi, A., Donahue, G., and Zaret, K.S. (2012) Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* 151, 994-1004.
151. Srivastava, R. and Ahn, S.H. (2015) Modifications of RNA polymerase II CTD: Connections to the histone code and cellular function. *Biotechnol Adv* 33, 856-872.
152. Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature* 403, 41-45.
153. Subramanian, V., Fields, P.A., and Boyer, L.A. (2015) H2A.Z: a molecular rheostat for transcriptional control. *F1000Prime Rep* 7, 01.
154. Takai, D. and Jones, P.A. (2004) Origins of bidirectional promoters: computational analyses of intergenic distance in the human genome. *Mol Biol Evol* 21, 463-467.
155. Talbert, P.B. and Henikoff, S. (2014) Environmental responses mediated by histone variants. *Trends Cell Biol* 24, 642-650.
156. Tan-Wong, S.M., Zaugg, J.B., Camblong, J., Xu, Z., Zhang, D.W., Mischo, H.E., . . . Proudfoot, N.J. (2012) Gene loops enhance transcriptional directionality. *Science* 338, 671-675.
157. Tark-Dame, M., Jerabek, H., Manders, E.M., Heermann, D.W., and van Driel, R. (2014) Depletion of the chromatin looping proteins CTCF and cohesin causes chromatin compaction: insight into chromatin folding by polymer modelling. *PLoS Comput Biol* 10, e1003877.
158. Temin, H.M. and Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 226, 1211-1213.
159. Terzi, N., Churchman, L.S., Vasiljeva, L., Weissman, J., and Buratowski, S. (2011) H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1-Nab3-Sen1 pathway. *Mol Cell Biol* 31, 3569-3583.
160. Thakar, A., Gupta, P., Ishibashi, T., Finn, R., Silva-Moreno, B., Uchiyama, S., . . . Zlatanova, J. (2009) H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. *Biochemistry* 48, 10852-10857.
161. Thatcher, T.H. and Gorovsky, M.A. (1994) Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. *Nucleic Acids Res* 22, 174-179.
162. Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P., and Myers, R.M. (2004) An abundance of bidirectional promoters in the human genome. *Genome Res* 14, 62-66.
163. van Arensbergen, J., van Steensel, B., and Bussemaker, H.J. (2014) In search of the determinants of enhancer-promoter interaction specificity. *Trends Cell Biol* 24, 695-702.

164. van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., . . . Morillon, A. (2011) XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* 475, 114-117.
165. Wal, M. and Pugh, B.F. (2012) Genome-wide mapping of nucleosome positions in yeast using high-resolution MNase ChIP-Seq. *Methods Enzymol* 513, 233-250.
166. Warner, J.R. (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* 24, 437-440.
167. Wasylyk, B. and Chambon, P. (1979) Transcription by eukaryotic RNA polymerases A and B of chromatin assembled in vitro. *Eur J Biochem* 98, 317-327.
168. Watanabe, S., Radman-Livaja, M., Rando, O.J., and Peterson, C.L. (2013) A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. *Science* 340, 195-199.
169. Weber, C.M., Ramachandran, S., and Henikoff, S. (2014) Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol Cell* 53, 819-830.
170. Wei, W., Pelechano, V., Jarvelin, A.I., and Steinmetz, L.M. (2011) Functional consequences of bidirectional promoters. *Trends Genet* 27, 267-276.
171. Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., . . . Davis, R.W. (1999) Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* 285, 901-906.
172. Wolfe, K.H. and Shields, D.C. (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387, 708-713.
173. Won, K.J., Choi, I., LeRoy, G., Zee, B.M., Sidoli, S., Gonzales-Cope, M., and Garcia, B.A. (2015) Proteogenomics analysis reveals specific genomic orientations of distal regulatory regions composed by non-canonical histone variants. *Epigenetics Chromatin* 8, 13.
174. Wong, M.M., Cox, L.K., and Chrivia, J.C. (2007) The chromatin remodeling protein, SRCAP, is critical for deposition of the histone variant H2A.Z at promoters. *J Biol Chem* 282, 26132-26139.
175. Xu, L.C., Thali, M., and Schaffner, W. (1991) Upstream box/TATA box order is the major determinant of the direction of transcription. *Nucleic Acids Res* 19, 6699-6704.
176. Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., . . . Steinmetz, L.M. (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033-1037.
177. Yan, C., Zhang, D., Raygoza Garay, J.A., Mwangi, M.M., and Bai, L. (2015) Decoupling of divergent gene regulation by sequence-specific DNA binding factors. *Nucleic Acids Res* 43, 7292-7305.
178. Yan, K., Arfat, Y., Li, D., Zhao, F., Chen, Z., Yin, C., . . . Qian, A. (2016) Structure Prediction: New Insights into Decrypting Long Noncoding RNAs. *Int J Mol Sci* 17.

179. Yang, M.Q. and Elnitski, L.L. (2008) Diversity of core promoter elements comprising human bidirectional promoters. *BMC Genomics* 9 Suppl 2, S3.
180. 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.
181. Zaragoza, D., Ghavidel, A., Heitman, J., and Schultz, M.C. (1998) Rapamycin induces the G0 program of transcriptional repression in yeast by interfering with the TOR signaling pathway. *Mol Cell Biol* 18, 4463-4470.
182. Zaret, K.S. and Carroll, J.S. (2011) Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* 25, 2227-2241.
183. Zhang, H., Roberts, D.N., and Cairns, B.R. (2005) Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* 123, 219-231.
184. Zhang, L., Ding, Q., Wang, P., and Wang, Z. (2013a) An upstream promoter element blocks the reverse transcription of the mouse insulin-degrading enzyme gene. *Biochem Biophys Res Commun* 430, 26-31.
185. Zhang, L., Wang, P., Ding, Q., and Wang, Z. (2013b) Transcriptional directionality of the human insulin-degrading enzyme promoter. *Mol Cell Biochem* 382, 237-242.
186. Zhou, L., Sun, K., Zhao, Y., Zhang, S., Wang, X., Li, Y., . . . Wang, H. (2015) Linc-YY1 promotes myogenic differentiation and muscle regeneration through an interaction with the transcription factor YY1. *Nat Commun* 6, 10026.
187. Zlatanova, J. and Thakar, A. (2008) H2A.Z: view from the top. *Structure* 16, 166-179.
188. Zofall, M., Fischer, T., Zhang, K., Zhou, M., Cui, B., Veenstra, T.D., and Grewal, S.I. (2009) Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. *Nature* 461, 419-422.