High-resolution studies of the chromatin and transcription landscape

Christopher M. Weber

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Steven Henikoff, Chair
Stephen Tapscott
Toshio Tsukiyama

Program Authorized to Offer Degree:
Molecular and Cellular Biology

University of Washington

Abstract

High-resolution studies of the chromatin and transcription landscape

Christopher M. Weber

Chair of Supervisory Committee:
Steven Henikoff Affiliate Professor
School of Medicine
Department of Genome Sciences

Transcription from genomic DNA is regulated in many ways and governs cellular identity, growth, and homeostasis. If transcription becomes deregulated, it can serve a critical role in the development of cancer and other human pathologies, however complete understanding of this process is lacking. In eukaryotes, transcriptional regulation involves a balance between repressive packaging of the genome into nucleosomes and enabling access to regulatory proteins as well as RNA polymerase II (RNAPII). Nucleosomes are strong physical barriers to transcription *in vitro* that cause RNAPII to backtrack and arrest. Yet, *in vivo* RNAPII must transcribe across many nucleosomes for every gene at a very high rate. How this happens and what mechanisms enable RNAPII transit through nucleosomes has long remained unclear. Here we show that the nucleosomes of active genes have a distinct histone composition which involves the replacement of both canonical H2A histones with histone variant H2A.Z. These homotypic H2A.Z nucleosomes show evidence of disruption during transcriptional elongation, suggesting that they are formed through transcription mediated turnover and then replacement with H2A.Z. Homotypic H2A.Z nucleosomes have unique physical properties and interact with distinct chromatin remodelers from canonical nucleosomes. Hence, we developed a single-nucleotide resolution approach to map RNAPII genome-wide and determine the nature of

transcription through these nucleosomes. We show that the entry site to the nucleosome is most refractory, contrary to existing models based on transcription *in vitro*, and that the first nucleosome from the transcription start site (+1) is a much larger barrier than downstream nucleosomes that causes RNAPII to backtrack. Nucleosome occupancy positively correlates with the magnitude of the barrier, however our results suggest that an evolved function of H2A.Z is to ease the inhibitory nature of nucleosomes on transcription. This helps to explain why H2A.Z is most enriched where the barrier is the largest and also why it is essential in development where transcriptional fidelity is critical.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am indebted to many people whom I have had the great privilege of learning from during my time as a graduate student. I would like to first thank my graduate advisor Steve Henikoff for his optimism, trust, and superb mentorship. I have learned from him how to think creatively about science, ask interesting questions, and be ambitious in solving them. For this I am extremely grateful. Over the years I have also had tremendous colleagues in the lab that have shaped me into a better scientist. I am especially grateful to Jorja Henikoff, Srinivas Ramachandran, Také Furuyama, Florian Steiner, Mary Gehring, Roger Deal, Pete Skene, Fan Yang, and Paul Talbert for their helpful comments, contributions, and discussions that helped to refine my thinking and research.

I would also like to thank my thesis committee for their support and guidance: Stephen Tapscott, Mark Groudine, Toshio Tsukiyama, and Richard Gardner. This research could not have been possible without funding from an NSF Graduate Research Fellowship and the Howard Hughes Medical Institute.

Finally, I would like to express deep gratitude to my wonderful wife Leah, who has always been extremely supportive of my research, especially when things were not working. In addition, I would like to acknowledge my family and friends, who have always been very supportive of my aspirations.

Chapter 1

Histone Variants: dynamic punctuation in transcription

Modified from a review published in Genes & Development

Abstract

Eukaryotic gene regulation involves a balance between packaging of the genome into nucleosomes and enabling access to regulatory proteins and RNA polymerase. Nucleosomes are integral components of gene regulation that restrict access to both regulatory sequences and the underlying template. Whereas canonical histones package the newly replicated genome, they can be replaced with histone variants that alter nucleosome structure, stability, dynamics, and ultimately DNA accessibility. Here we consider how histone variants and their interacting partners are involved in transcriptional regulation through the creation of unique chromatin states.

Introduction

The basic unit of chromatin is the nucleosome, consisting of DNA wrapped around a core of histone proteins. Nucleosomes likely evolved to protect and compact increasingly large eukaryotic genomes (Malik and Henikoff, 2003). As a consequence however, nucleosomes also restrict access to cellular components such as DNA-binding transcription factors and RNA polymerase (Li et al., 2007). Accordingly, distinct mechanisms have evolved to influence the dynamic competition between nucleosomes and DNA-binding transcription factors in addition to orchestrating RNA polymerase II (RNAPII) translocation across a nucleosomal template. This dynamic mode of regulation is mediated in a number of ways including: post-translational

modification of histones, altering the position or eviction of nucleosomes by ATP-dependent chromatin remodelers, and replacement of canonical histones with histone variants. Canonical histones (H2A, H2B, H3, and H4) are deposited in a replication-coupled manner to package the newly replicated genome. In contrast, histone variants are expressed throughout the cell cycle and replace canonical histones or take their place when nucleosomes are evicted (see (Malik and Henikoff, 2003; Talbert and Henikoff, 2010) for reviews on evolutionary conservation). Histone variants have distinct amino acid sequences that can influence both the physical properties of the nucleosome and nucleosome dynamics. These properties are especially important during transcription where histone variants shape the chromatin landscape of cis-regulatory and coding regions, in support of specific transcription programs.

Here we consider core variants of H2A and H3 that are implicated in transcription (Table 1-1). We discuss the mechanisms responsible for shaping the histone variant landscape, focusing on recent genome-wide mapping studies of these variants, their chaperones, RNA polymerase II, and nucleosome dynamics. We review how these chromatin landscapes and deposition pathways influence the dynamic interplay between nucleosome occupancy, regulatory DNA-binding proteins, and ultimately RNA polymerase II elongation across nucleosomes. Our ultimate focus is on how histone variants create distinct chromatin landscapes with different dynamics and how this influences gene regulation.

Histone Variant	Presence
H2A.Z	Universal
macroH2A	
macroH2A	Invertebrates
macroH2A.1.1 / 2	Vertebrates
macroH2A.2	Vertebrates
H2A.B	Vertebrates
H2A.Lap1	Mouse
H2A.Bbd	Human
H3.3	Universal (only H3 in fungi)

Table 1-1
Histone variants discussed implicated in transcriptional regulation and the lineages where they are present.

Nucleosome organization and dynamics

Each nucleosome wraps ~147 bp of DNA 1.7 turns around an octamer consisting of two of each of H2A, H2B, H3, and H4. At the center of the DNA wrap an $(H3/H4)_2$ tetramer is formed due to a strong four-helix bundle interaction between the two H3 proteins (Luger et al., 1997). Interacting with the $(H3/H4)_2$ tetramer are two heterodimers of H2A/H2B, which dock at the DNA entry and exit sites through the H2A C-terminus docking domain (Figure 1-1A). Additionally, the two H2A histones interact through their L1 loop and H2B interacts with H4 through a weak four-helix bundle.

Nucleosomes are energetically stable, however they can turn over in vivo (Figure 1-1B). H2A and H2B turn over much faster than H3 and H4 in both genic and intergenic regions (Jamai et al., 2007; Kimura and Cook, 2001). H2A/H2B turnover occurs due to weaker intranucleosomal contacts and because the dimers dock at DNA entry and exit sites, which are prone to transiently unwrap (Li et al., 2005b). Active processes such as RNA polymerase transit drives unwrapping, which increases dimer turnover and DNA exposure (Sheinin et al., 2013). A

functional consequence of both transient unwrapping and dimer loss is exposure of DNA to binding by regulatory proteins.

In order for the $(H3/H4)_2$ tetramer to turn over, the nucleosome must be almost completely unwrapped, a process that can occur many times at a given location during interphase. In yeast and Drosophila cells, nucleosome turnover measured by new H3 incorporation correlates with transcription (Deal et al., 2010; Dion et al., 2007). In Drosophila cells, nucleosome turnover is also high over cis-regulatory regions, consistent with a competition between nucleosomes and transcription factors for occupancy of these sites (Deal et al., 2010; Mito et al., 2007). A consequence of the dynamic nature of nucleosomes is erasure of post-translational modifications on histones. Thus, modulation of DNA accessibility through regulated nucleosome turnover can perpetuate gene expression states.

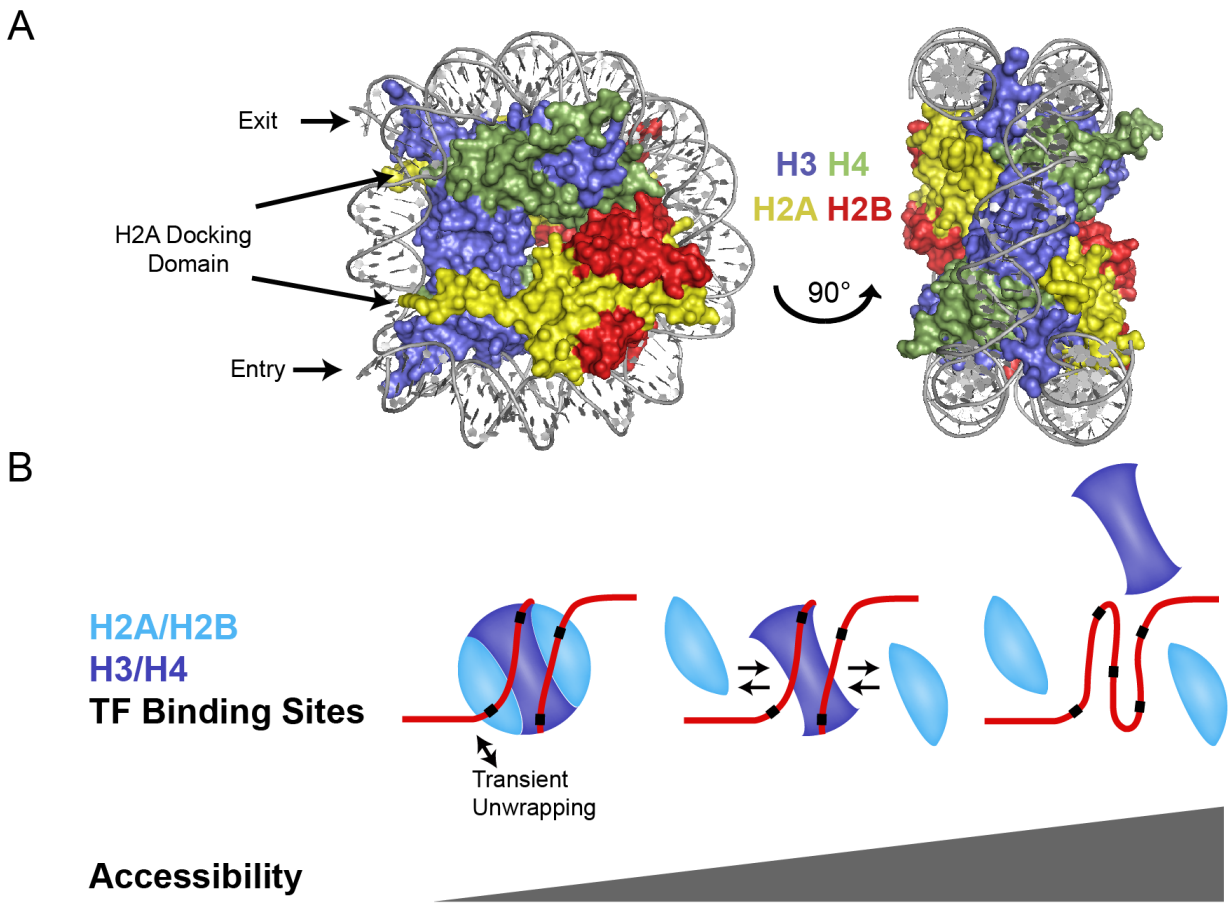


Figure 1-1. Nucleosome organization and dynamics
A) Structure of nucleosome core particle, showing that $(H3/H4)_2$ is at the center of the DNA wrap, with two dimers of H2A/H2B docked at the edges, near the DNA entry and exit locations. B) Nucleosomes modulate access to transcription factor (TF) binding sites. Transient DNA unwrapping exposes transcription factor binding sites. As nucleosomes unwrap H2A/H2B dimers can be lost, exposing more DNA, and when the nucleosome is completely unwrapped $(H3/H4)_2$ is lost, DNA is completely exposed.

Replication-Independent deposition of variants punctuates the chromatin landscape

H2A.Z alters nucleosome properties

H2A variants are the most diverse, perhaps reflective of relaxed structural constraint within the nucleosome. One such variant, H2A.Z, arose once early in eukaryotic evolution and has remained distinct from H2A ever since (Talbert and Henikoff, 2010). At the amino acid level, H2A.Z is only ~60% identical to H2A within species but is relatively conserved between species, and is essential in metazoans (Zlatanova and Thakar, 2008). Remarkably, the structure of the H2A.Z nucleosome is quite similar to H2A, however there are key structural differences (Suto et al., 2000). On the surface, H2A.Z has an extended acidic patch, which stimulates remodeling activity with the ISWI ATP-dependent remodeler (Goldman et al., 2010). Within the core, the L1 loop is structurally distinct, and in the docking domain with H3/H4 a glutamine to glycine substitution in H2A.Z compromises three hydrogen bonds, which is predicted to weaken the interaction. Despite these structural differences, the change in the stability of the particle is subtle, with contrasting results reported. Overall, the consensus is that H2A.Z slightly stabilizes *in vitro* and destabilizes *in vivo* (Bonisch and Hake, 2012; Zlatanova and Thakar, 2008). This discrepancy might be attributable to post-translational modifications in vivo, where H2A.Z is acetylated at active genes (Bruce et al., 2005; Valdes-Mora et al., 2012), or to differences in DNA sequence, or the fact that H2A.Z nucleosomes can be hybrid in vivo, either heterotypic (Z/A) or homotypic (Z/Z) (Luk et al., 2010; Viens et al., 2006; Weber et al., 2010).

H2A.Z deposition and a futile cycle

In yeast, H2A.Z can be bound by the general H2A/H2B chaperone Nap1 or Chz1which preferentially binds H2A.Z over H2A (Luk et al., 2007). These chaperones provide a source of H2A.Z for the Swr1 remodeling complex, which exchanges H2A.Z for H2A (Mizuguchi et al., 2004). Some metazoans contain two Swr1 orthologs that organize into at least two distinct complexes, P400/TIP60 and SRCAP, which like Swr1 catalyze the exchange reaction. While these complexes share some components, there are many differences (Billon and Cote, 2012). For example, a human H2A.Z specific chaperone, ANP32E, was recently characterized as part of P400/TIP60 but not the SRCAP complex (Obri et al., 2014). Intriguingly, the ANP32 family has many members of uncharacterized function, where at least one other, ANP32B, has been shown to be an H3/H4 chaperone (Tochio et al., 2010). It is likely that many other histone chaperones remain to be discovered.

H2A.Z comprises ~15% of total H2A and is distributed throughout the genome non-randomly in both euchromatin and heterochromatin, where it is monoubiquitinated (Sarcinella et al., 2007), however, it has remained unclear how H2A.Z becomes enriched at particular sites in the genome. High-resolution ChIP (chromatin immunoprecipitation) and biochemistry of the yeast Swr1 complex revealed that Swr1 preferentially acts at nucleosome depleted regions (NDRs) that are >50-70bp and requires the Swc2 subunit to bind DNA (Ranjan et al., 2013; Yen et al., 2013). NDRs are predominately located at the promoter region of active genes and are characterized by two well-positioned flanking nucleosomes. Certain transcription factors and chromatin remodelers are required for NDR establishment and maintenance, but H2A.Z and Swr1 are dispensable (Hartley and Madhani, 2009; Whitehouse et al., 2007). Whereas recruitment to the NDR might be sufficient to explain H2A.Z enrichment at nucleosomes that flank promoters, H2A.Z is also enriched to some extent in gene bodies of all eukaryotes studied.

Additionally, NDR recruitment does not explain how some organisms, including Arabidopsis (Zilberman et al., 2008) and Drosophila (Mavrich et al., 2008), lack upstream H2A.Z nucleosomes (Figure 1-2A).

In metazoans, H2A.Z enrichment correlates with expression level. One attractive possibility is that dimer loss followed by replacement with H2A.Z contributes to genic enrichment patterns and low-level incorporation genome-wide. In support of this possibility, homotypic Z/Z nucleosomes are enriched over active genes (Nekrasov et al., 2012; Weber et al., 2010), which could be the result of transcription-mediated dimer loss and then opportunistic replacement with H2A.Z, which is produced throughout the cell cycle. Consistent with transcription-coupled replacement, the presence of upstream H2A.Z nucleosomes seen in some organisms correlates with bidirectional transcription in yeast and mammals (Core et al., 2008; Xu et al., 2009), whereas upstream H2A.Z enrichment is not seen in Arabidopsis and Drosophila. DNA methylation and H2A.Z are anti-correlated in plants and animals (Conerly et al., 2010; Zemach et al., 2010) and in Arabidopsis there is mutual antagonism (Zilberman et al., 2008), suggesting that epigenetic factors also contribute to H2A.Z localization.

In mammalian cells preferential H2A.Z localization also is seen at a subset of gene promoters associated with transcription factor binding (Gallant-Behm et al., 2012; Gevry et al., 2007). Additionally, work in embryonic stem cells (ESCs) has shown that H2A.Z preferentially localizes to promoters of silent genes occupied by the Polycomb Repressive Complex 2 (PRC2) (Creyghton et al., 2008; Illingworth et al., 2012). H2A.Z is also enriched at enhancers in ESCs and facilitates the binding of PRC2 and H3K4me3 and H3K27me3 modifications (Hu et al., 2013). Also in ESCs, H2A.Z increases nucleosome accessibility at FoxA2 binding sites (Li et al., 2012) and transcription factor accessibility (Hu et al., 2013). Despite these intriguing

observations, it remains unclear how H2A.Z becomes enriched at developmentally regulated regions.

H2AZ can be actively removed from chromatin. For example, the recently described ANP32E chaperone specifically removes H2A.Z from nucleosomes, including those at cis-regulatory sites (Obri et al., 2014). The Ino80 complex catalyzes replacement of H2A.Z with canonical H2A (Papamichos-Chronakis et al., 2011), just the reverse of the Swr1 complex, thus completing a futile cycle when both complexes act successively on the same nucleosome. In an Ino80 mutant, H2A.Z is globally mislocalized, suggesting that Ino80 clears out H2A.Z that is spuriously incorporated. Ino80 is also found at yeast NDRs, where high-resolution ChIP suggests that Nhp10, les5, and Arp8 subunits are important for binding, similar to the Swc2 subunit of the Swr1 complex (Yen et al., 2013). In addition, Swr1 can carry out the reverse reaction when H3K56 is acetylated in yeast (Watanabe et al., 2013). However, when Swc2 is present, this activity is inhibited suggesting that this subunit functions as a lock to prevent further remodeling. In yeast, H3K56 is acetylated during replication-coupled nucleosome assembly and is present on ~30% of total histone H3 (Xu et al., 2005), whereas it marks less than 1% in human cells (Xie et al., 2009). Considering its paucity in metazoans, it remains unclear as to how general a role H3K56ac plays in removal of H2A.Z.

Other H2A variants: life on the edge

Unlike H2A.Z, which is nearly universal across eukaryotes, other replication-independent H2A variants implicated in transcription have evolved only in animals. For example, H2A.B, first described as H2A.Bbd (Barr body deficient) has only been found in mammals. H2A.B is only ~50% identical to H2A and is rapidly evolving (Ishibashi et al., 2010). The C-terminus of

H2A.B is 19 amino acids shorter than that of H2A, reducing the tail and part of the docking domain. H2A.B lacks an acidic patch on its surface and is also referred to as H2A.Lap (lacks acidic patch) (Soboleva et al., 2012). These modifications to the structure substantially alter the physical properties of H2A.B containing nucleosomes. For example, human H2A.B nucleosome arrays inhibit the formation of compact chromatin fibers (Zhou et al., 2007). Mouse H2A.B can form partially compacted arrays due to a single aspartate residue in the acidic patch, which is not found in human (Soboleva et al., 2012). Consistent with an integral role for the docking domain in modulating nucleosome stability, H2A.B nucleosomes are less stable, protect ~30bp less DNA against micrococcal nuclease digestion, and exchange much faster by FRAP than canonical H2A (Bao et al., 2004; Bonisch and Hake, 2012; Doyen et al., 2006b; Gautier et al., 2004; Tolstorukov et al., 2012). These effects are largely attributable to the docking domain because adding the H2A C-terminal tail, including the docking domain onto H2A.B, partially reverses the instability (Doyen et al., 2006b).

Currently, it remains unclear whether there are specific chaperones or mechanisms for H2A.B deposition. However, the general chaperone NAP-1 can efficiently assemble and disassemble H2A.B dimers in vitro (Okuwaki et al., 2005). In HeLa cells, ectopically expressed H2A.B localizes over gene bodies and correlates with expression level (Tolstorukov et al., 2012). Similarly, in ESCs endogenous H2A.B is enriched over the body of actively transcribed genes (Chen et al., 2014). In contrast, in mouse testis where H2A.B is especially abundant, it is enriched over gene promoters and lowly over gene bodies (Soboleva et al., 2012). It remains unclear why H2A.B localization is different, nonetheless, H2A.B deposition results in a distinct chromatin landscape that is destabilized and less compact.

Another H2A variant implicated in transcription, macroH2A, is distinct from all other histones in that it contains a non-histone globular (macro) domain. The macro domain on the C-terminus is connected through an unstructured linker to a histone fold domain that is ~60% identical to canonical H2A, resulting in a histone protein that is ~3x the size of H2A (Chakravarthy et al., 2005). Macro-domain containing proteins are found in many organisms, however macroH2A is restricted to vertebrates and a few invertebrates (Talbert and Henikoff, 2010). Macro domains are known to bind metabolites of NAD^+ including Poly(ADP-ribose) and have distinct biological roles including transcriptional regulation (Han et al., 2011). In vertebrates, there are three macroH2A isoforms: macroH2A1.1, macroH2A1.2 and macroH2A2. The first two are splice isoforms from a single gene, whereas a separate gene encodes the latter. However, only macroH2A.1.1 is capable of binding metabolites of NAD+. In contrast to both H2A.Z and H2A.B, macroH2A preferentially forms heterotypic nucleosomes in vitro, whereas H2A.Z shows no preference (Chakravarthy et al., 2004; Chakravarthy and Luger, 2006). macroH2A has a higher salt-dependent stability than H2A, where four residue changes in the L1 loop are most important (Abbott et al., 2005; Chakravarthy and Luger, 2006).

In vivo, macroH2A is enriched on the transcriptionally inactivated female X chromosome, senescence-associated heterochromatic foci (SAHF), and large transcriptionally silent domains (Costanzi and Pehrson, 1998; Gamble et al., 2010; Tolstorukov et al., 2012; Zhang et al., 2005b). macroH2A localization to the inactive X is disrupted when Xist RNA is deleted, suggesting a role for this essential cis-regulatory RNA in macroH2A recruitment (Csankovszki et al., 1999). Both the histone and non-histone segments of macroH2A are sufficient for targeting to the inactive X (Chadwick et al., 2001; Nusinow et al., 2007). In SAHF, macroH2A deposition is promoted by histone regulator A (HIRA), a chaperone responsible for

histone variant H3.3 incorporation at genes, and the nucleosome assembly and disassembly factor Asf1 (Zhang et al., 2005b). Rather than directly deposit macroH2A, these chaperones might help to clear the way for macroH2A deposition. Consistent with higher stability and a generally repressive role, macroH2A appears to impair transcription factor binding and has been suggested to impair remodeling by SWI/SNF and ISWI (Angelov et al., 2003). Although more recent results suggest that macroH2A is an excellent substrate for remodeling by these remodelers, macroH2A reduces recruitment of the SWI/SNF remodeler (Chang et al., 2008). macroH2A also associates with the SWI/SNF-family DNA translocase ATRX (α-thalassemia/MR, X-linked), which like Ino80 for H2A.Z, negatively regulates macroH2A deposition (Figure 1-2B) (Ratnakumar et al., 2012). Despite identification of these partners for macroH2A, it remains unclear how it is localized into specific genomic regions.

H3.3: filling gaps and more

When a nucleosome is lost independent of replication, the $(H3/H4)_2$ tetramer is replaced by the H3.3 variant and its H4 partner. Most eukaryotes express canonical H3 for replication-coupled deposition and the replication-independent H3.3 variant, however some eukaryotes, such as fungi, express only the H3.3 type (Malik and Henikoff, 2003). In metazoans, H3.3 differs from H3 by only 4-5 amino acids (Filipescu et al., 2013). Three of these differences are found within the core histone fold domain and specify the alternative deposition pathways (Ahmad and Henikoff, 2002). The fraction of the genome occupied by H3.3 is variable, depending largely on dilution by canonical H3 during replication. For example, H3.3 comprises ~ 90% of the histone 3 in terminally differentiated neurons (Pina and Suau, 1987) however, in dividing cells H3.3 comprises only ~20% (McKittrick et al., 2004). H3.3 incorporation is largely opportunistic,

occurring when DNA is exposed at dynamic regions, such as gene promoters, the body of active genes, and cis-regulatory elements (Mito et al., 2007; Ray-Gallet et al., 2011; Schneiderman et al., 2012). H3.3 was also found to be enriched over a subset of repressed genes in mammalian ESCs, which exhibit lower dynamics, suggesting an expanded role for H3.3 beyond simply filling gaps in active chromatin (Goldberg et al., 2010). H3.3 enrichment over genes, both active and inactive, depends upon the HIRA complex (Tagami et al., 2004). In addition, HIRA independent mechanisms of H3.3 incorporation have been described (Banaszynski et al., 2013). For example, Daxx, has been identified as a novel H3.3-specific chaperone, which together with the SWI/SNF family remodeler, Atrx, is responsible for incorporation at telomeres and pericentric heterochromatin (Drane et al., 2010; Goldberg et al., 2010). A single methionine to glycine substitution at position 90 in H3.3 appears to be a dominant contributor to the specificity of H3.3 interaction with the Daxx chaperone (Elsasser et al., 2012). In Drosophila Daxx and Dek have been shown to deposit H3.3 in regulatory elements (Sawatsubashi et al., 2010).

From a structural perspective H3 occupies the center of the nucleosome and so it might not be surprising that the sequence of an H3 variant would be more constrained than is seen for variants of H2A. However, this constraint is not observed in the centromere-specific H3 variant, cenH3, which only shares ~50-60% identity with H3 within the histone fold domain (Malik and Henikoff, 2003). Considering that the alterations to H3.3 are subtle, it is not surprising that no destabilization attributable to H3.3 nucleosomes has been detected in vitro (Chen et al., 2013b; Thakar et al., 2009). However, analysis of chromatin from chicken cells found that H3.3 containing nucleosomes are more sensitive to salt-dependent disruption and that H3.3/H2A.Z double variant nucleosomes were most unstable (Jin and Felsenfeld, 2007). This effect was independent of acetylation, a modification associated with destabilization of nucleosomes,

suggesting that the effect is intrinsic or is due to incorporation at active regions of the genome. In support of the latter explanation, H3.3/H2A.Z nucleosomes are enriched over regulatory elements and NDRs which are frequently disrupted (Jin et al., 2009). However, nucleosome turnover and HIRA binding to chromatin are reduced after H3.3 depletion (Banaszynski et al., 2013). HIRA was recently shown to directly interact with transcription factors, and the Brg1 chromatin remodeling complex (Pchelintsev et al., 2013). These results overall suggest that incorporation of H3.3 promotes a hyperdynamic state through its interaction partners within the nucleus.

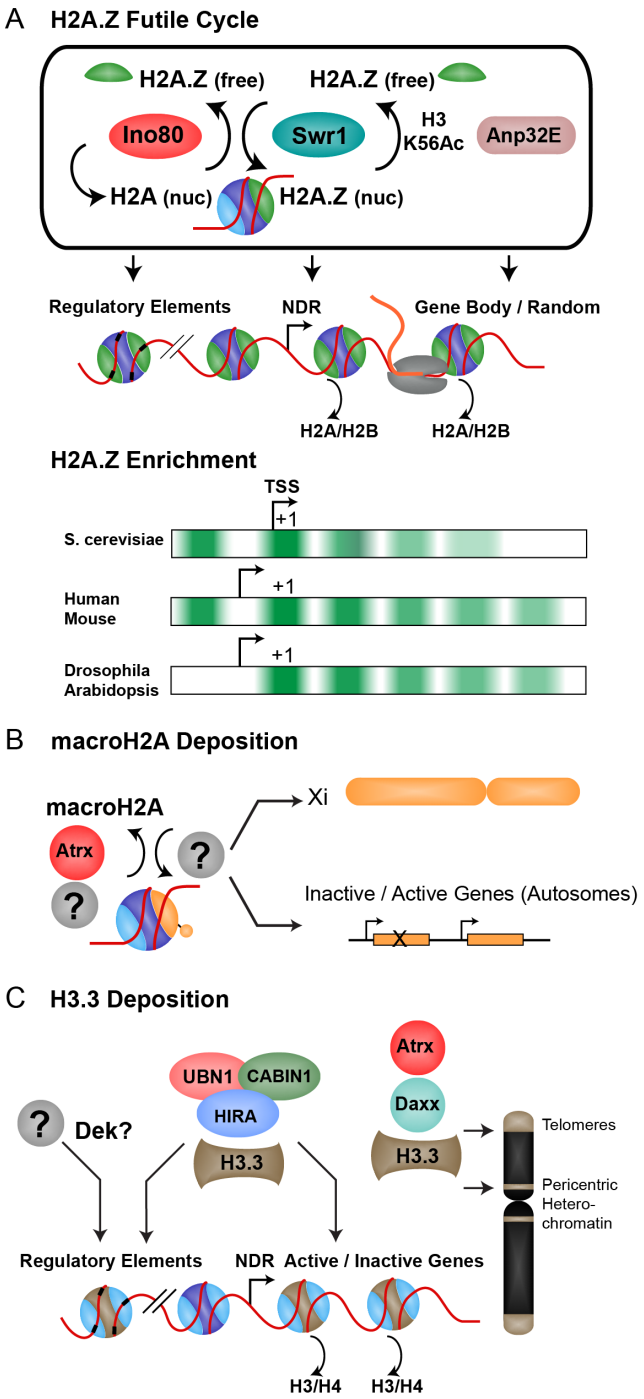


Figure 1-2. Histone variant deposition pathways
A) (Top) The H2A.Z futile cycle of deposition by Swr1 or orthologous complexes and then removal by Ino80, Anp32E, or Swr1 when H3K56 is acetylated. (Bottom) H2A.Z enrichment patterns and promoter architecture differences in different organisms. B) macroH2A is deposited in the inactive X (Xi) and inactive/active genes on autosomes. C) H3.3 Deposition is mediated by the Hira complex at genes and regulatory elements, whereas Atrx and Daxx mediates H3.3 incorporation at telomeres and pericentric heterochromatin. It is not clear how H3.3 is enriched at some cis-regulatory elements.

Histone variants in transcriptional regulation

H2A.Z: The positive, negative, and the unknown

A role for H2A.Z in transcription was initially proposed over 30 years ago with the observation that Tetrahymena H2A.Z is present in the transcriptionally active macronucleus but not in the inactive micronucleus (Allis et al., 1980). Initial studies investigating the effect of H2A.Z on transcription were conducted in yeast where it was shown to antagonize telomeric silencing, interact with activators, and to help recruit RNAPII (Adam et al., 2001; Santisteban et al., 2000) (Meneghini et al., 2003). Upon activation, H2A.Z is lost from 5' ends of genes, suggesting that it poises genes for activation by enabling access to the promoter/TSS. In human cells, H2A.Z is exchanged prior to RNAPII loading, likely as a consequence of promoter remodeling, where it has a role in RNAPII recruitment (Hardy et al., 2009). It is possible that the difference in the action of H2A.Z between yeast and metazoans stems from the fact that in yeast there is a nucleosome over the transcription start site that must be removed for RNAPII to load (Figure 1-2A) (Rhee and Pugh, 2012). Regardless of the nucleosome architectural at promoters, H2A.Z consistently plays an activating role (Figure 1-3).

In addition to roles in activation and initiation, H2A.Z has been found to promote elongation (Santisteban et al., 2011). To explore the molecular basis for this role we developed a single base resolution method to map the position of RNAPII through the 3' end of nascent transcripts in Drosophila (3'NT). We found an anti-correlation between H2A.Z occupancy and RNAPII stalling as it transcribes across nucleosomes genome-wide (Weber et al., 2014). When H2A.Z levels were reduced both directly and by impairing Swr1 activity, the nucleosome barrier to RNAPII increased. Our results favor a model whereby H2A.Z/H2B dimers are more easily

lost when the nucleosome is unwrapped, aiding RNAPII transcriptional elongation. It is also possible that elongation factors such as the FACT complex have increased activity with H2A.Z nucleosomes to ease the barrier or that acetylation 'unlocks' H2A.Z nucleosomes by destabilizing them. For example, in yeast H2A.Z and Spt16, a component of the FACT complex, are synthetic lethal (Biswas et al., 2006). Also, H2A.Z is hyperacetylated only when genes are expressed (Bruce et al., 2005; Valdes-Mora et al., 2012) and acetylation is required to achieve proper transcript levels (Halley et al., 2010; Valdes-Mora et al., 2012).

Despite a generally positive role for H2A.Z in transcription, negative effects have also been reported. For example, yeast h2a.zΔ mutants derepress silencing at the HMR locus (Dhillon and Kamakaka, 2000) and whole-genome transcriptome analysis identified both positive and negative changes (Meneghini et al., 2003). In mammals, H2A.Z negatively regulates p21 and targets of the p63 transcription factor, and occupies the promoters of genes that are silenced during mitosis (Gallant-Behm et al., 2012; Gevry et al., 2007; Kelly et al., 2010). Intriguingly, the h2a.zΔ phenotype, which extends beyond transcription defects, is partially suppressed by also mutating integral components of the Swr1 complex (Halley et al., 2010; Morillo-Huesca et al., 2010). This suggests that Swr1 activity in the absence of H2A.Z causes 'mischief' by removing H2A but is unable to replace it. In the absence of both H2A.Z and Swr1 more yeast genes are downregulated than upregulated (Morillo-Huesca et al., 2010). However, in Arabidopsis h2a.zΔ swr1Δ (pie1-5) mutants are worse off than either individual mutant, suggesting non-redundant functions for Swr1 (Coleman-Derr and Zilberman, 2012). In Arabidopsis mutants that cannot incorporate H2A.Z, there is also global misregulation of transcription with many genes upregulated and many downregulated. Interestingly, in Arabidopsis H2A.Z is lost at elevated temperatures independent of transcription, and the transcriptome of the incorporation mutant

resembles that of temperature-shifted plants (Kumar and Wigge, 2010). Thus, in both yeast and plants, where H2A.Z mutants are viable, loss of the H2A.Z deposition pathway results in global effects on transcription.

How might H2A.Z function to exert both positive and negative effects on transcription? An attractive explanation is that H2A.Z facilitates binding of both activating and repressive complexes by keeping regions of the genome accessible. Support for this model comes from work in ESCs and during differentiation, which has shown that H2A.Z facilitates binding of PRC2, MLL, and transcription factors (Creyghton et al., 2008; Hu et al., 2013; Li et al., 2012). In this context, H2A.Z deposition increases the fraction of the genome that is nuclease hypersensitive and decreases nucleosome occupancy at enhancers (Hu et al., 2013). Interestingly, nucleosome depletion at transcription factor binding sites is dependent upon SWI/SNF and Ino80 chromatin remodeling complexes (Li et al., 2012). This suggests that a function of the futile cycle of deposition and removal might be to modulate accessibility to various regulatory proteins. It was previously suggested that H2A.Z regulates nucleosome positioning around promoters, which could influence accessibility of cis-regulatory regions (Guillemette et al., 2005; Marques et al., 2010). However, we and others failed to detect changes in nucleosome positioning following H2A.Z knockdown or deletion (Hartley and Madhani, 2009; Li et al., 2005a; Weber et al., 2014). Although there might be intrinsic effects of H2A.Z on translational or rotational positioning in yeast (Albert et al., 2007), it is also possible that these attributes are due to enriching for promoter flanking nucleosomes, having little to do with the properties of H2A.Z.

Overall, H2A.Z functions to support transcriptional activation and elongation, which helps to explain its enrichment near promoters and over the coding region of all genes in

eukaryotes. Future work will determine which components are involved in H2A.Z dynamics and how these accessible regions are regulated both positively and negatively. This is especially relevant in mammalian cells, which deposit H2A.Z through P400 and SRCAP complexes, whose independent functions are not yet clear. That H2A.Z functions to facilitate access to both repressive and active regulatory complexes explains the vexing dual nature of H2A.Z in transcription, and helps to explain why H2A.Z is essential development.

H2A.B

Whereas the role of H2A.B in transcription has not been as thoroughly investigated as that of H2A.Z, recent experiments have provided interesting results. H2A.B knockdown in HeLa cells resulted in substantial changes in gene expression, with more genes downregulated than upregulated (Tolstorukov et al., 2012). In mouse ESCs, H2A.B knockdown resulted in a more modest effect on gene expression, however most genes were downregulated. Included in the differentially expressed set are a few imprinted genes, where it was shown that H2A.B facilitates elongation over the differentially methylated region of the gene (Chen et al., 2014). Consistent with a role for H2A.B in transcriptional elongation, H2A.B has also been shown to associate with components of the spliceosome, and upon knockdown RNA splicing is less efficient (Tolstorukov et al., 2012). In mouse testis, H2A.B is associated with promoter regions, as discussed above, although it is not yet known whether it has any specific effect on expression (Soboleva et al., 2012). Overall, H2A.B appears to promote transcriptional elongation likely as a consequence of creating less stable nucleosomes that are more easily disrupted upon interaction with RNA polymerase (Figure 1-3).

macroH2A

macroH2A localization on the inactivation female X chromosome, silent SAHF foci, and large transcriptionally silent domains (Costanzi and Pehrson, 1998; Zhang et al., 2005b) (Gamble et al., 2010; Tolstorukov et al., 2012) suggests a role in transcriptional repression, although macroH2A is not required for X chromosome inactivation (Changolkar et al., 2007). Genome-wide studies in human NT2 cells have shown that macroH2A is enriched at developmentally regulated genes, including overlap with the PRC2 complex. Nevertheless, it also represses transcription in vitro (Doyen et al., 2006a) and in vivo represses IL8 transcription in a human B cell line and endogenous retroviruses in mice (Agelopoulos and Thanos, 2006; Changolkar et al., 2008). Knockdown of macroH2A increased the sensitivity of genes in the HOXA cluster to retinoic acid, further suggesting that it acts to repress transcription (Buschbeck et al., 2009). In the human breast cancer cell line, MCF-7, most genes with macroH2A enrichment are not expressed, however macroH2A depletion did not cause their upregulation. Somewhat surprisingly, activating roles have also been reported for macroH2A, such as serum starvation induced genes (Gamble et al., 2010) (Figure 1-3). Currently, the mechanism for this dual role of macroH2A remains unclear. It is possible that some of these discrepancies can be explained by macroH2A1.1 inhibition of PARP-1 (Ouararhni et al., 2006) or Poly(ADP-ribose) modification, both of which are known to be involved in transcriptional regulation.

H3.3: A dynamic memory

The function of H3.3 in transcription remains somewhat unclear. Studies in Tetrahymena showed that H3.3 is not essential for transcription or viability (Cui et al., 2006). In adult Drosophila males, the loss of both H3.3 genes results in partial lethality and mostly affects

highly expressed genes, with more genes upregulated than downregulated overall (Sakai et al., 2009). However, constitutive expression of H3 largely rescued these effects, suggesting that transcription differences were a consequence of nucleosome depletion and not specifically H3.3. In H3.3 depleted or HIRA -/- ESCs, a minor fraction of genes showed differences in transcript levels, some of which are developmentally regulated, however there is a much larger effect when H3.3 is depleted in partially differentiated mouse embryonic fibroblasts (Banaszynski et al., 2013; Goldman et al., 2010).

Recent evidence suggests that there are unique roles for H3.3 and its interacting partners in transcription. For example, H3.3 is important for early gene activation in cell lines and during myogenic differentiation (Placek et al., 2009; Tamura et al., 2009; Yang et al., 2011). In Xenopus, HIRA-mediated deposition of H3.3 is required for the transcriptional memory of active genes after somatic cell transfer into enucleated eggs (Jullien et al., 2012; Ng and Gurdon, 2008). H3.3 has also been shown to prime genes for later activation after genotoxic stress (Adam et al., 2013). Although the molecular basis for transcriptional memory is unknown, it is intriguing that H3.3/H4 tetramers at human enhancer elements split during replication (Huang et al., 2013). H3.3 deposition might facilitate transcription factor binding by keeping these regions of the genome accessible (Figure 1-3). Experiments in ESCs also support this general role, where H3.3 knockdown compromised PRC2 binding as well as H3K27 trimethylation at ‘bivalent promoters’ (Banaszynski et al., 2013). A similar role has been ascribed to H2A.Z and it is thus conceivable that there is some crosstalk or that increased H2A.Z dynamics influence H3.3 deposition and dynamics.

Perspective

Over the past few years there has been a growing appreciation for histone variants in transcriptional regulation. From the evidence described in this review, chromatin mediated gene regulation acts primarily through modulation of nucleosome dynamics and access to the underlying DNA. Although variants are structurally distinct from their canonical counterparts, replacement of a canonical histone with a variant is not the only way to alter stability and dynamics of nucleosomes. An emerging theme is that variants function as an ensemble, coordinately modifying nucleosome properties and interacting with an expanding catalog of other factors within the nucleus. Differences in physical properties of variants and interactions with trans-acting factors result in dynamic punctuation of chromatin that profoundly influences accessibility of the genome and ultimately transcriptional regulation. This is perhaps most evident during development in metazoans, which involves global changes in chromatin organization and transcriptional programs. A major area of interest for the future will be the characterization of metazoan deposition complexes, where we anticipate many context-dependent roles, and the emergence of many additional players.

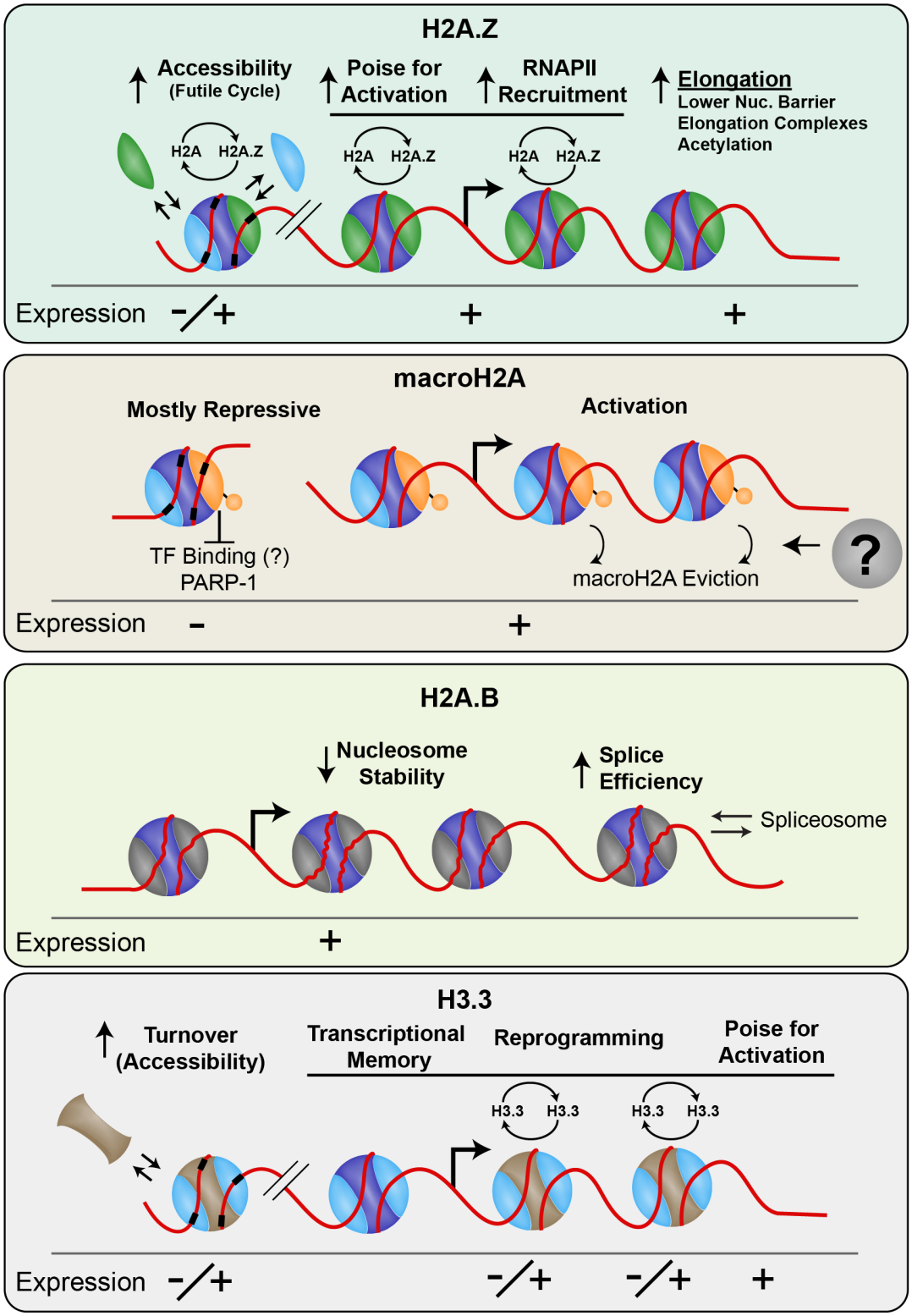


Figure 1-3. Models to explain general role of histone variants and their deposition pathways on transcriptional regulation.
Plus or minus expression level denotes the effect of the variant on transcriptional output.

Chapter 2

H2A.Z nucleosomes enriched over active genes are homotypic

Modified from a paper published in Nature Structural and Molecular Biology

Abstract

Nucleosomes that contain the histone variant H2A.Z are enriched around transcriptional start sites, but the mechanistic basis for enrichment is unknown. A single octameric nucleosome can contain two H2A.Z histones (homotypic) or one H2A.Z and one canonical H2A (heterotypic). We generated high-resolution maps of homotypic and heterotypic Drosophila H2A.Z (H2Av) nucleosomes. Although homotypic and heterotypic H2A.Z nucleosomes map throughout most of the genome, homotypic nucleosomes are enriched and heterotypic nucleosomes are depleted downstream of active promoters and intron/exon junctions. The distribution of homotypic H2A.Z nucleosomes resembles that of classical active chromatin and shows evidence of disruption during transcriptional elongation. Both homotypic H2A.Z nucleosomes and classical active chromatin are depleted downstream of paused polymerases. Our results suggest that H2A.Z enrichment patterns result from intrinsic structural differences between heterotypic and homotypic H2A.Z nucleosomes following disruption during transcriptional elongation.

Introduction

The eukaryotic genome is packaged into nucleosome core particles that consist of ~147 bp of DNA wrapped around an octamer of four core histone proteins: H2A, H2B, H3, and H4, which must be modified to facilitate access to DNA for active processes. Alterations to nucleosomes include remodeling, post-translational modifications, and replacement of resident

histones with histone variants. Whereas the major canonical histones are expressed during S phase and package the newly synthesized DNA into nucleosomes, replacement histone variants are expressed from distinct genes throughout the cell cycle and have important functional roles(Talbert and Henikoff, 2010). For example, the universal histone variant H2A.Z is a minor variant that comprises only ~5-10% of major H2A(Palmer et al., 1980; Wu et al., 1982; Zlatanova and Thakar, 2008)} yet H2A.Z is essential for viability in many organisms, including *Tetrahymena thermophila*(Liu et al., 1996), *Xenopus laevis*(Ridgway et al., 2004) *Mus musculus*(Faast et al., 2001) and *Drosophila melanogaster*(van Daal and Elgin, 1992). Despite extensive characterization, the role of H2A.Z in gene regulation has remained elusive.

Exchange of major H2A for variant H2A.Z within the nucleosome is mediated by the Swr1 ATPase complex(Kobor et al., 2004; Krogan et al., 2003; Mizuguchi et al., 2004). Studies in *Saccharomyces cerevisiae* using H2A.Z chromatin immunoprecipitation (ChIP) coupled with microarray technology (ChIP-chip) found that H2A.Z nucleosomes are distributed throughout the genome but are enriched at gene promoters(Zlatanova and Thakar, 2008). In *Drosophila* embryos, H2A.Z localizes in the proximity of ~85% of transcription start sites (TSS)(Mavrich et al., 2008). Many studies have looked more closely at the localization of H2A.Z at promoters, and a conserved profile across eukaryotes has emerged. H2A.Z localizes to the strongly positioned +1 nucleosome downstream from the TSS and a few positioned nucleosomes further downstream, but is relatively depleted over gene bodies(Barski et al., 2007; Lantermann et al., 2010; Mavrich et al., 2008; Zilberman et al., 2008). In both *S. cerevisiae* and mammals, H2A.Z is also enriched upstream of TSSs, although this enrichment is not seen in *Drosophila*, *Arabidopsis,* or *Schizosaccharomyces pombe*. The basis for differential H2A.Z localization is unknown.

H2A.Z arose early in eukaryotic evolution and has remained distinct from H2A ever since(Talbert and Henikoff, 2010). Relative to H2A, H2A.Z has an extended acidic patch, which alters the nucleosome surface to interactions with other proteins, and a different docking domain, which interacts with H3 within the core particle. H2A.Z and H2A also have drastically different L1 loops, which comprise the only interaction surface between the pair of H2A-H2B dimers that flank the $(H3\text{-}H4)_2$ core of the nucleosome(Suto et al., 2000). Whereas both the H2A-H2A and the H2A.Z-H2A.Z interaction surfaces are tightly interlocked, the H2A.Z-H2A.Z interaction is more extensive. Also, it does not appear that the nucleosome core particle can accommodate an H2A-H2A.Z interaction without a structural clash. Nevertheless, subsequent studies have shown that heterotypic nucleosomes containing both H2A and H2A.Z form *in vitro*(Chakravarthy et al., 2004), and *in vivo* are more abundant than homotypic (H2A.Z-H2A.Z-containing) nucleosomes(Viens et al., 2006). *In vitro* studies have indicated that these different structural forms have different inherent properties. For example, reconstituted homotypic H2A.Z nucleosomes are relatively resistant to disruption by high ionic strength solutions(Li et al., 1993; Park et al., 2004; Thambirajah et al., 2006) and are refractory to RNA polymerase passage(Thakar et al., 2010). Furthermore, after acetylation, reconstituted homotypic H2A.Z nucleosomes are more stable than either heterotypic H2A.Z or canonical H2A nucleosomes(Ishibashi et al., 2009).

We wondered whether these different structural properties between canonical, heterotypic and homotypic nucleosomes might contribute to H2A.Z localization and function *in vivo*. Accordingly, we generated high resolution, genome-wide maps of heterotypic and homotypic H2A.Z nucleosomes in Drosophila S2 cells. We find that heterotypic and homotypic H2A.Z nucleosomes are distributed almost indistinguishably genome-wide. However, homotypic H2A.Z

nucleosomes are highly enriched just downstream of TSSs of active genes, where heterotypic H2A.Z nucleosomes are relatively depleted, and are enriched downstream of intron/exon junctions. Homotypic H2A.Z nucleosomes also correspond closely to low-salt-soluble nucleosomes, and are depleted from genes regulated by RNA Polymerase II (Pol II) pausing, where both are deficient. Our findings lead to a structure-based dynamic model that can help explain both H2A.Z localization and molecular function.

Results

Distribution of homotypic and heterotypic H2Av nucleosomes

To separate heterotypic and homotypic H2A.Z nucleosomes, we used sequential affinity purification with two different tags on H2A-type histones. *Drosophila* has a single replication-independent H2A variant, H2Av, which functions both as H2A.Z and H2A.X(Malik and Henikoff, 2003). *Drosophila* S2 cell lines that express biotin-tagged H2A (Biotag-H2A) or H2Av (Biotag-H2Av) histones were infected with baculovirus expressing FLAG-tagged H2Av driven by the inducible *Drosophila* metallothionein promoter (Figure 2-2). We prepared native micrococcal nuclease digested (MNased) chromatin from these two baculovirus-infected lines using a low-salt EDTA shearing protocol, which maximally solubilizes chromatin under minimally disruptive conditions(Jin and Felsenfeld, 2007). We then fixed soluble nucleosomes consisting of mostly mononucleosomes with formaldehyde to ensure integrity of labile nucleosomes through subsequent purification steps. We isolated DNA from the tagged nucleosomes bound to streptavidin beads (Figure 2-1a-b), and sequenced the purified nucleosomal DNA using the paired-end mode of the Illumina Genome Analyzer (Figure 2-3). We also sequenced a fraction of the input from one heterotypic and one homotypic H2Av

purification experiment, which consisted of mostly mononucleosomes. The sequenced input DNA displayed extensive genome-wide coverage (Figure 2-1). Surprisingly, profiles from both heterotypic and homotypic H2Av nucleosomes also showed broad genome-wide distributions. Although both heterotypic and homotypic H2Av nucleosomes are nearly ubiquitous throughout the genome, quantitative differences were evident in comparing input, heterotypic, and homotypic H2Av nucleosomes. The close correspondence between biological replicates in each case indicates that these differences are robust. We also analyzed transposon families and found major differences in nucleosome density, as indicated by their very different transposon landscapes for input, heterotypic and homotypic nucleosomes (Figure 2-4). However, we detected only slight differences between input, heterotypic, and homotypic H2Av nucleosomes within a family (Figure 2-4b). Therefore, H2Av nucleosomes of both types are broadly distributed throughout the *Drosophila* genome.

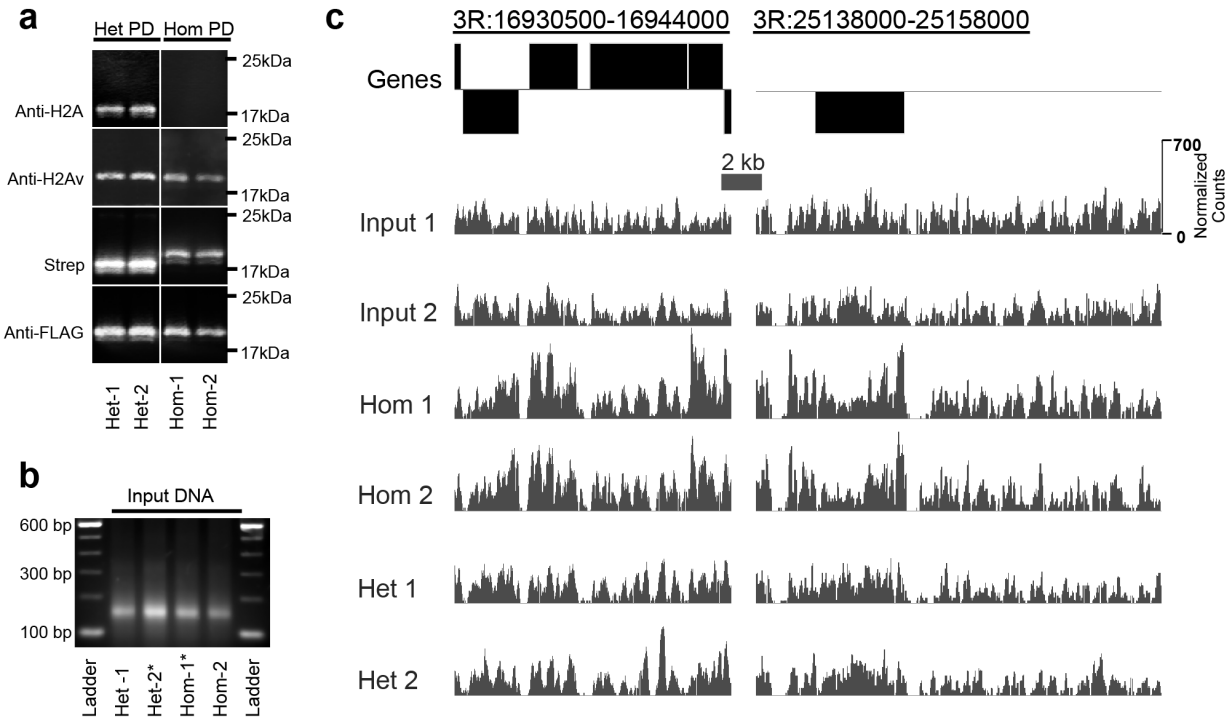


Figure 2-1. Broad distribution of homotypic and heterotypic H2Av nucleosomes.

(a) Western analyses of two replicates shows that sequential affinity-purified heterotypic and homotypic nucleosomes contain the expected composition of tagged histones and that endogenous epitopes are preserved. From the heterotypic purification (Het PD), antibodies (or streptavidin) recognize endogenous epitopes and tags from Biotag-H2A (16.9kDa) and FLAG-H2Av (18kDa). From the homotypic purification (Hom PD), antibodies [anti-H2A (Upstate 07-146), anti-H2Av(Madigan et al., 2002) or anti-FLAG (Sigma, F3165)] or streptavidin recognize endogenous epitopes and tags from Biotag-H2Av (18.5kDa) and FLAG-H2Av (H2Av). Proteins extracted from streptavidin beads with affinity-purified nucleosomes were resolved on 18% (w/v) SDS-PAGE gels and transferred to nitrocellulose. (b) Agarose gel showing that input DNA (1µg) comprises mostly mononucleosomal fragments for two replicates. Asterisks mark input samples that were sequenced. (c) Normalized counts in 10-bp intervals from representative regions of chromosome 3R, showing two biological replicates for all DNA sizes for input, homotypic, and heterotypic sequenced libraries.

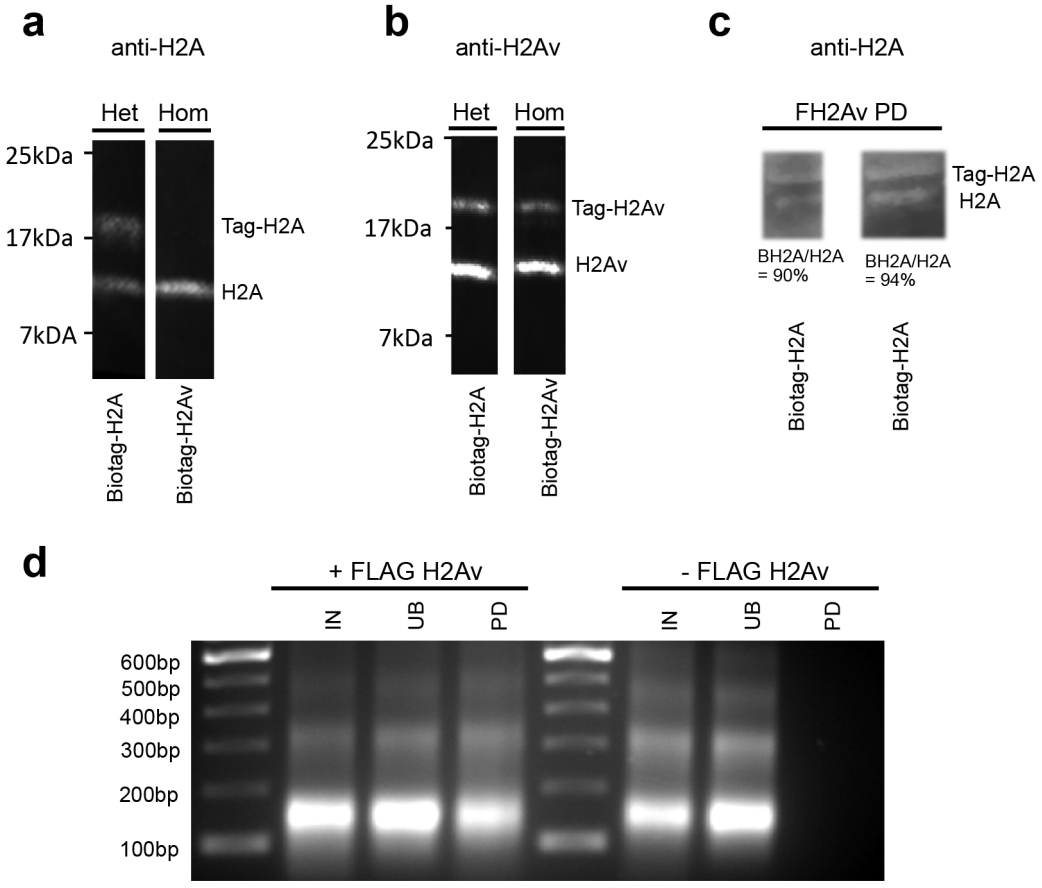
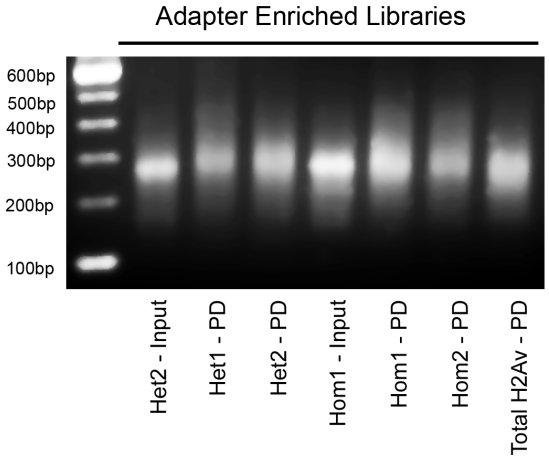


Figure 2-2. Analysis of chromatin from input to affinity purification and single pulldown. (a-c) Western analysis of chromatin input to heterotypic (Het) and homotypic (Hom) affinity purifications and single FLAG-H2Av pulldowns from Biotag-H2A transformed cells probed with antibodies against H2A and H2Av epitopes. (a) Biotag-H2A (16.9 kDa) protein level is roughly equivalent to endogenous H2A (13.4 kDa) in Biotag-H2A expressing cells. As expected there is only endogenous H2A in the Biotag-H2Av cell line. (b) FLAG-H2Av (18.0

kDa) levels in the Biotag-H2A cell line are lower than endogenous H2Av (15.0 kDa). In Biotag-H2Av transformed cells, the FLAG-H2Av and Biotag-H2Av (18.5 kDa) co-migrate, and their combined protein level is less than endogenous but is comparable to the FLAGH2Av protein level alone in the Biotag-H2A cell line. (c) H2A levels are nearly equivalent to Biotag-H2A in single FLAG-H2Av pulldown from biological replicates. Tagged versus endogenous ratios were quantified using LI-COR Odyssey software. (d) Mock pulldown does not recover DNA. S2 cells were seeded the same as other experiments, but one flask was infected with FLAG-H2Av baculovirus and the other was not. Chromatin was prepared the same way as for all other experiments as well as the FLAG IP. Following affinity purification, the beads were washed and 0.5µg from input (IN), unbound (UB), and pulldown (PD) was loaded onto a 2% Agarose gel. In the case of the mock transduced PD, the capacity of the well was loaded with the sample.

a



b

Sample	Total Reads	Paired-End Reads
Homotypic Input 1	8726228	4146288
Homotypic PD 1	7694251	4128614
Homotypic PD 2	5880032	3350819
Heterotypic Input 2	8802996	4448350
Heterotypic PD 1	8214518	4218733
Heterotypic PD 2	7518220	3672361
Total H2Av (Single PD)	9193255	4522575
80mM low-salt soluble	4849566	N/A

Figure 2-3. Illumina library preparation.

(a) An illustrative example is shown. After linker ligation and PCR amplification, samples were electrophoresed to verify the size distribution. Primers and adapters contribute 119 bp, so that the gradually diminishing smear of fragments below the mononucleosome band is consistent with the read distribution shown in Figure 2-5a. (b) Read counts for this experiment. Only uniquely mapped paired-end reads with paired alignment score > 0 were counted.

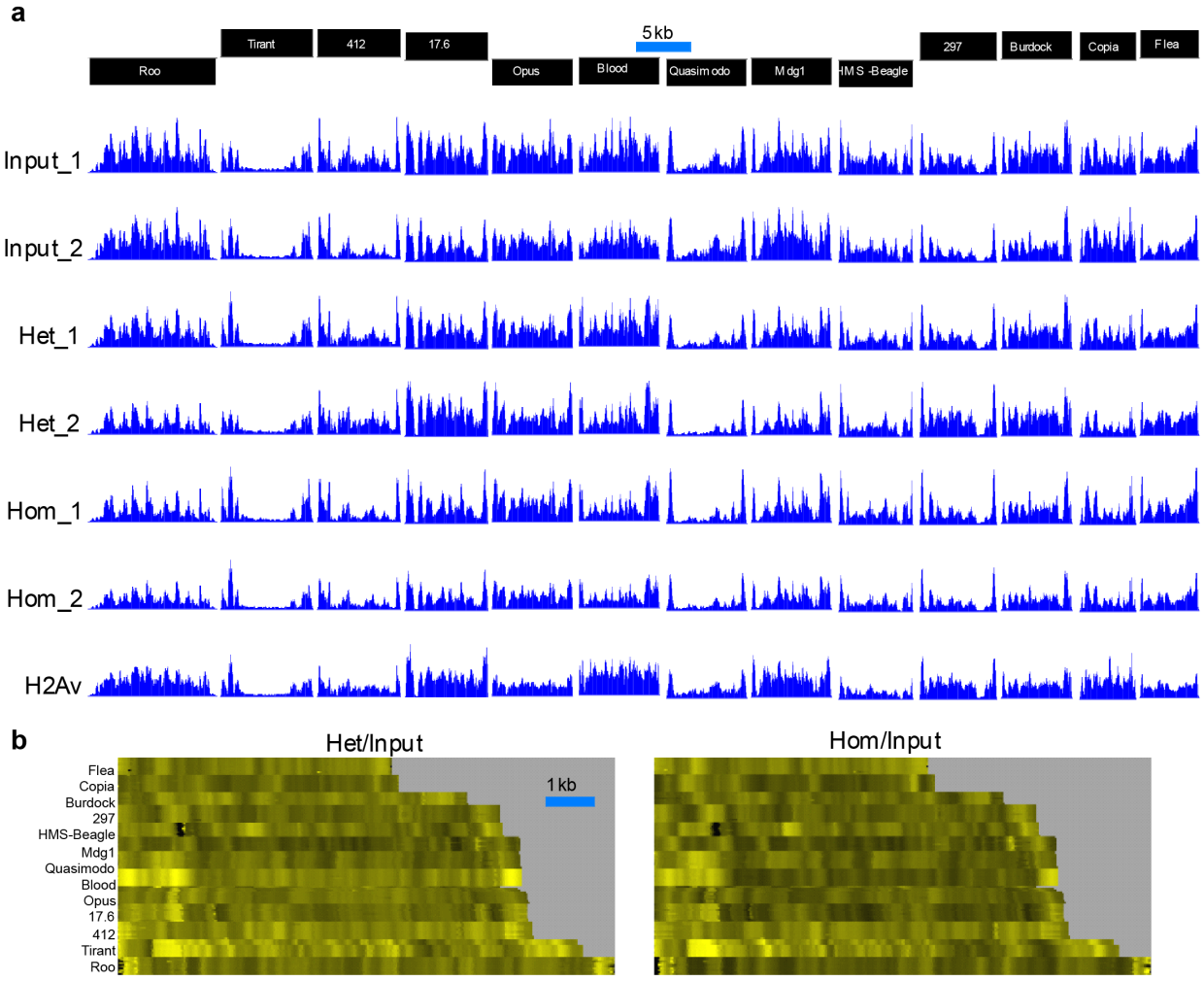


Figure 2-4. Profiles at LTR retrotransposon families are similar regardless of nucleosome composition.
(a) Counts from individual datasets for input, heterotypic, homotypic, and total H2Av at longest representative transposon from each family. (b) Heat map stack of up to 10 aligned elements representing each of the 13 families ordered by length and profiled for mean of replicates for heterotypic/input and homotypic/input, which display very similar heat maps.

Homotypic H2Av nucleosomes are enriched downstream of TSSs

The sequencing of paired-end libraries precisely maps particles that are protected from MNase digestion, so that the integrity of the fragments used to generate libraries can be ascertained by bulk analysis of the fragment sizes that result. For each sequenced library, we observed a broad size distribution ranging from 35-bp to 333-bp (Figure 2-5**)**. The input distribution was characterized by a sharp maximum at 147-bp, as expected for fully protected mononucleosomal DNA flanked by exposed DNA that is subject to endonucleolytic nicking, double-strand cleavage and end-nibbling, which is characteristic of single-strand-specific nucleases such as MNase(Desai and Shankar, 2003). In addition, maxima were observed at approximately 55, 90, and 130 bp, with a trail of longer fragments greater than 180 bp. Similar distributions were seen for the affinity purified fractions, except that the 147-bp maxima were less prominent, and maxima also appeared around 170 bp. Because the large majority of fragments were mononucleosome or smaller in size, and no single fragment of the ~30 million sequenced was larger than 333 bp, we can be confident that our nucleosome purifications are not contaminated by tagged histones on neighboring nucleosomes.

Based on the distribution of fragment sizes and peaks, we binned the fragments into six size class intervals: 55, 90, 130, 147, 170, and >180-bp (Figure 2-5, top). For each class, we divided all 10,997 genes for which both 5' and 3' ends are known into five quintiles based on expression levels and profiled the reads at the ends of the genes. A surprising finding was that input nucleosomal fragments in the 55-bp size class (35-75 bp interval) peaked ~20-bp upstream of the TSSs for genes in the first, second and third quintiles, which represent the transcriptionally active genes. This interval roughly corresponds to the average footprint of Pol II(Buratowski et al., 1989; Selby et al., 1997). However, at least some of the peak may be nucleosomal, because

fragments in this size class were also enriched after H2Av affinity purification. It is likely that any nucleosomes yielding such short DNA fragments are sufficiently disrupted for MNase to cleave extensively within the particle.

For all five size classes larger than 55 bp, a striking difference between heterotypic and homotypic H2Av nucleosomes was evident. Homotypic H2Av was strongly enriched immediately downstream of the TSS of actively transcribed genes, where heterotypic H2Av was relatively depleted (Figure 2-5b and Figures 2-6 & 2-7). Peaks of homotypic H2Av nucleosomes displayed phasing from +1 to +4 with gradual dampening. The first homotypic H2Av expression quintile showed depletion compared to the second and third, which were similarly enriched. The fact that all five larger homotypic H2Av size classes independently showed the same profile, with only minor differences in peak resolution between quintiles, implies that these differences are robust. Furthermore, all five homotypic H2Av classes were strongly enriched compared to heterotypic H2Av. As a control we also sequenced the input fractions and, as expected, observed intermediate enrichments for the five larger size classes (Figure 2-5 and Figure 2-6). We have previously shown that H2A-containing nucleosomes are strongly depleted downstream of TSSs in a pattern that is approximately inverse to that of H2Av(Henikoff et al., 2009), which, when taken together with the global abundance of H2Av(Palmer et al., 1980), implies that homotypic H2Av nucleosomes predominate over the first few nucleosomes downstream of the TSS of active genes (Figure 2-8).

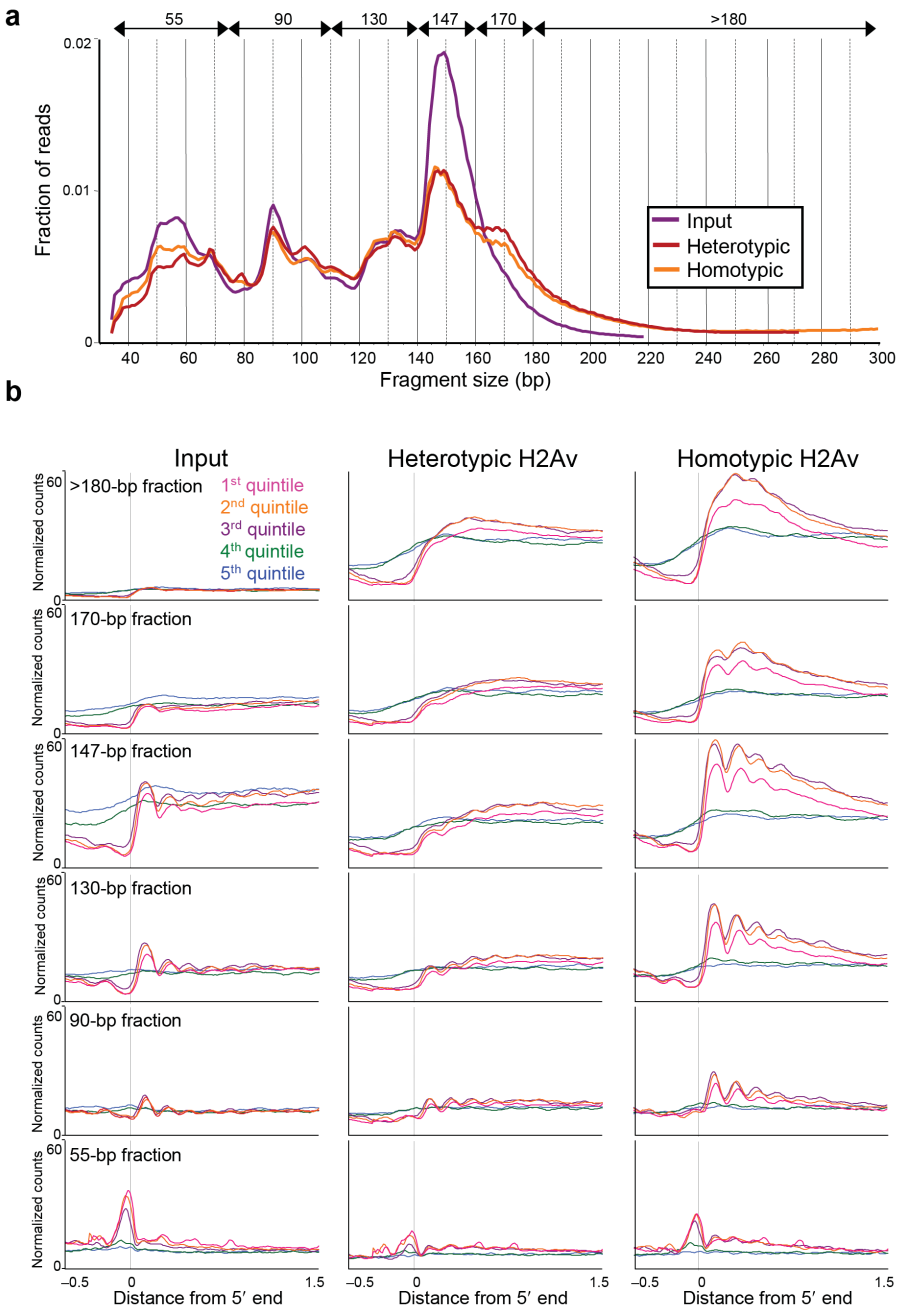


Figure 2-5. Homotypic H2Av nucleosomes are enriched downstream of gene promoters. (a) Histogram showing the length distribution of mapped paired-end reads from the mean of two biological replicates for input, heterotypic, and homotypic H2Av purifications. The size range of binned reads is indicated at the top of the graph by arrows and the corresponding size class is labeled. (b) Average profiles for input, heterotypic, and homotypic H2Av purifications for the six size-class intervals: 55-bp (35–75bp), 90-bp (76–110bp), 130-bp (111–140bp), 147-bp (141–160bp), 170-bp (161–180bp), >180-bp (181–333bp). Genes were divided into quintiles based on expression level for all 10,997 fully annotated genes, aligned at their TSSs and averaged within 10-bp bins. Averaging was truncated at 5' or 3' ends of neighboring genes. The *y*-axis refers to mapped paired-end counts. See also Figure 2-6 for 3’ ends as well.

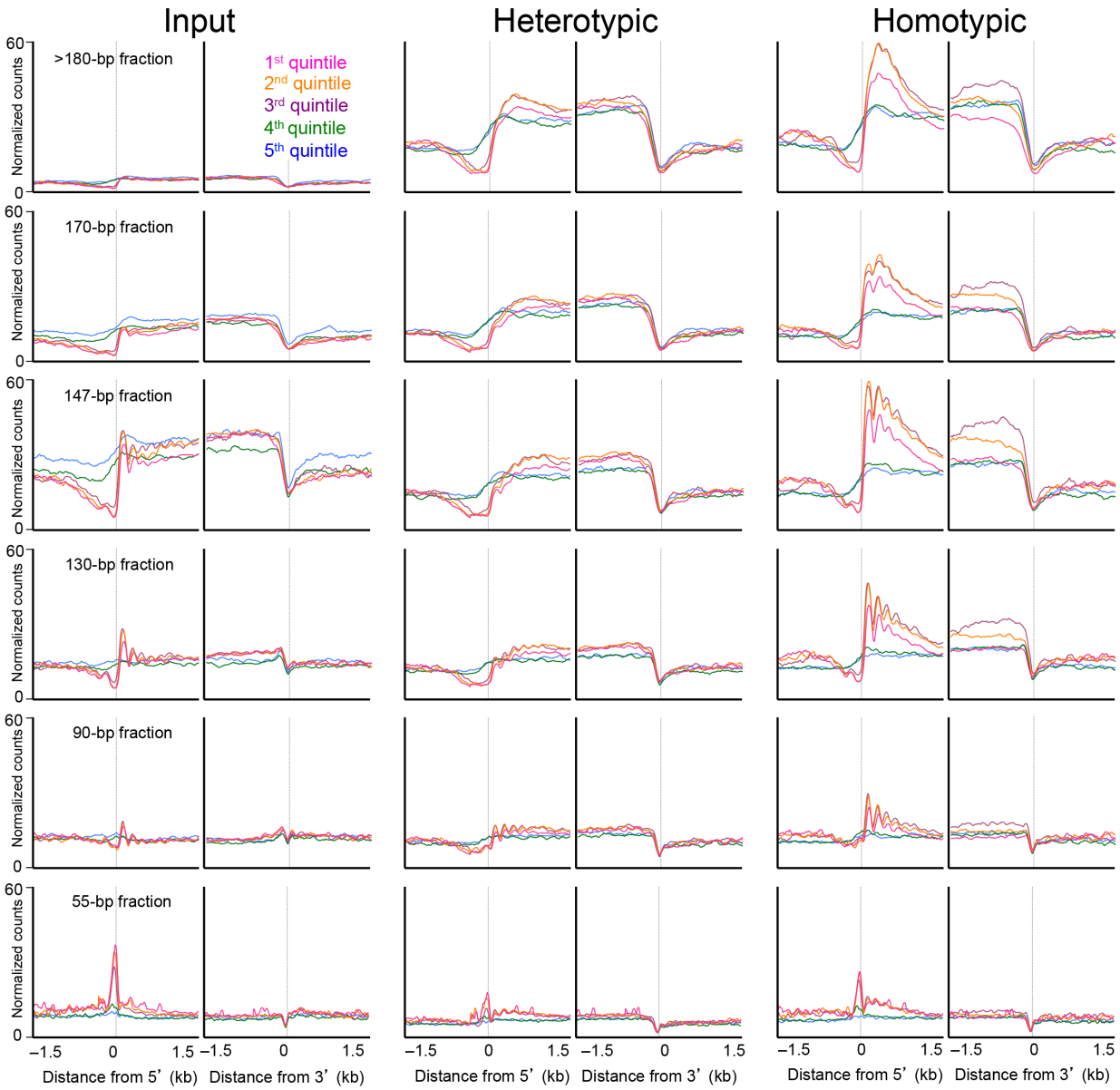


Figure 2-6. Size-class profiles around 5' and 3' ends of genes.
For each size class the mean of replicates from input, heterotypic, and homotypic is displayed +/- 1.5kb from the 5' and 3'ends of genes, divided into quintiles based on expression level. See the legend to Figure 2-5b for details.

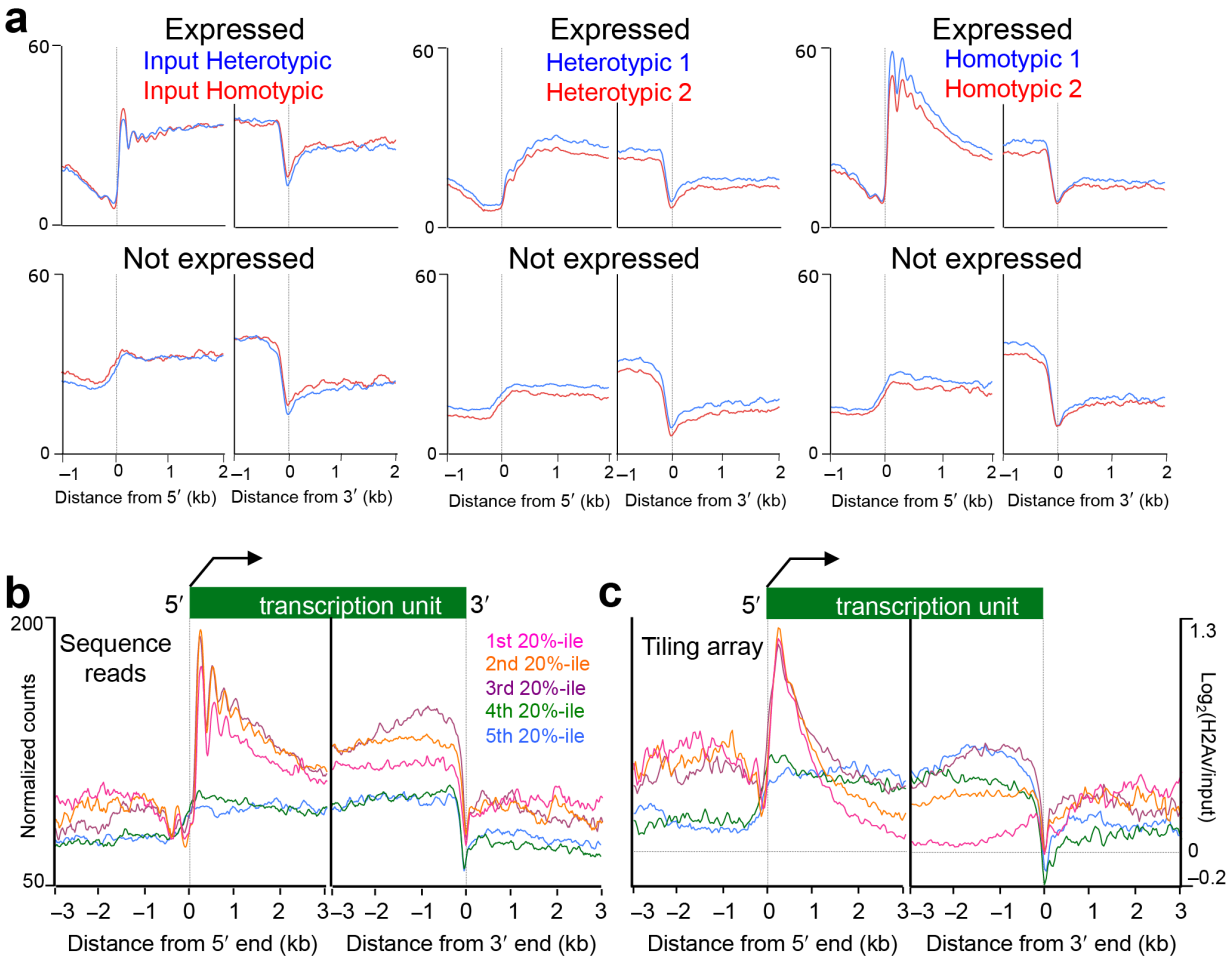


Figure 2-7. Comparisons of biological replicates and of profiles using different readout platforms.

(a) Individual profiles from the 147-bp (141-160 bp) size class for input, heterotypic and homotypic purifications profiled -1 to +2 kb from the 5' and 3' ends of active and inactive genes. (b) A Flag-H2Av sample was divided following MNase treatment and one aliquot subjected to tiling microarray analysis using the NimbleGen system and one aliquot subjected to sequencing using the Illumina system. b) Normalized paired-end counts within 10-bp intervals for the 147-bp size interval. c) Log-ratios of H2Av/input integrated over 25-bp intervals (GEO:GSM535514; Bryson et al. Fly 4:10.4161/fly. 4.3.12177). The relatively higher intergenic H2Av signal for the tiling array data is attributable to the lower intergenic nucleosome density (e.g., Figure 2-5b input), which largely cancels out in the log-ratio calculation. Slope differences might be due to the linear (b) versus logarithmic (c) scale differences using the two platforms.

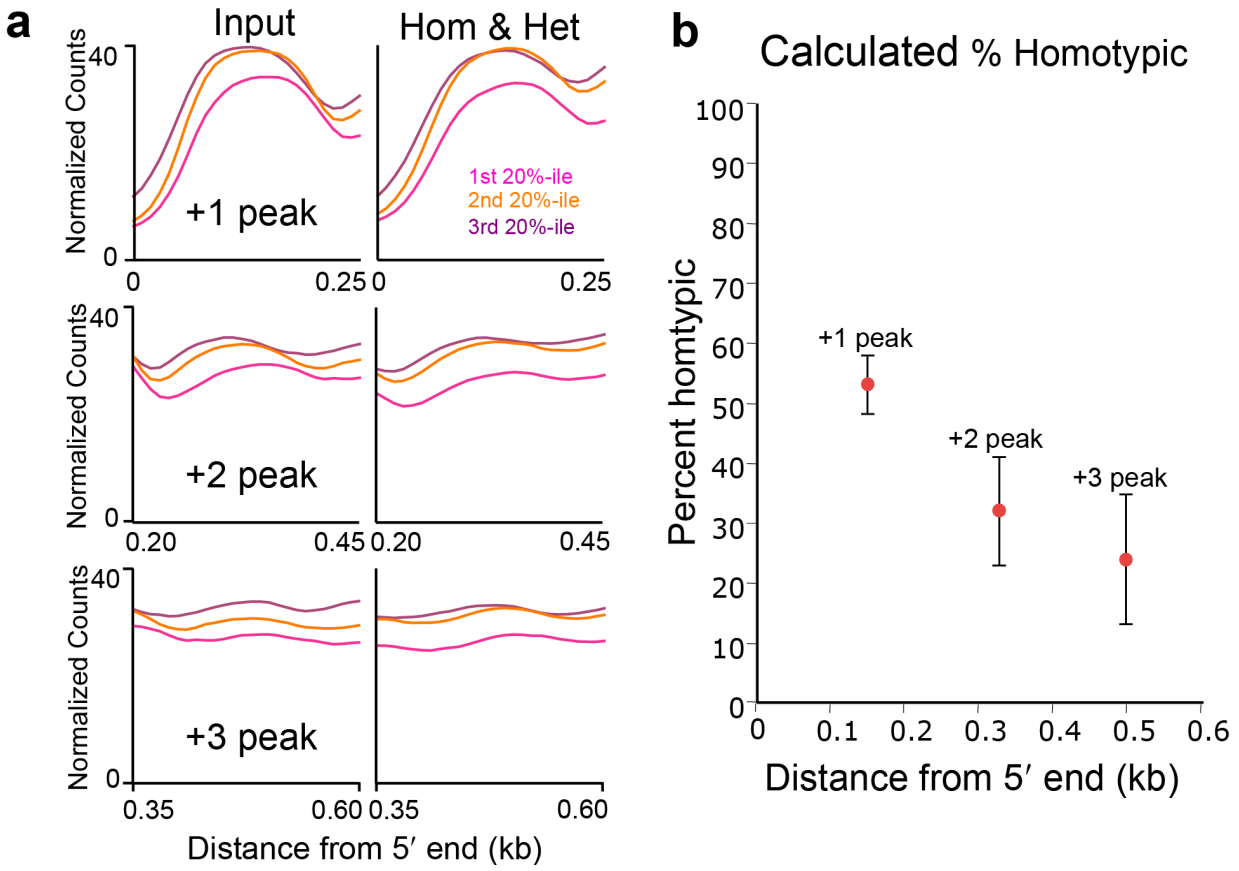


Figure 2-8. Evidence that homotypic H2Av nucleosomes predominate in the +1 phased position of active genes.

H2Av comprises 10% of the total H2A-type histone (Palmer et al., PNAS 77:2671, 1980), so that random incorporation would be expected to yield 1% [(10%)2] homotypic H2Av. To estimate the fraction of homotypic nucleosomes at phased peaks, we first note that H2A/H2A nucleosomes are strongly depleted around 5' regions of active genes (Henikoff et al. (Genome Res.19:460, 2009), and assume for simplicity that these peaks are entirely homotypic H2Av and heterotypic nucleosomes. We reasoned that the height of each peak observed in the input fraction should be equal to the sum of the heights of homotypic + heterotypic peaks, in units of normalized counts within a size class. a) Individual peaks are displayed using the mean number of normalized counts in the 147 size class at each of top 3 expression quintiles for the 1st, 2nd and 3rd peaks of homotypic H2Av (m) and heterotypic (t). We calculated their relative ratios such that their sum would equal the sum of normalized counts for input (I), I.e., I = x*m + (1-x)*t, where x is the fraction of homotypic nucleosomes. This yielded average values of x = 0.530, 0.315 and 0.234 for the +1, +2 and +3 phased nucleosome peaks, respectively, plotted as peak heights from the top 3 quintiles for two experiments (b). The close correspondences between the quintiles observed for measured versus calculated input provides evidence that our comparison is valid, and the fact that similar results were obtained for two experiments indicates that our estimates are robust. The trend is consistent with there being 1/10th as many homotypic H2Av as heterotypic nucleosomes genome-wide, consistent with random expectation, and with the results of Viens et al. (MCB 26:5325, 2006), who found that heterotypic H2A.Z is in large excess over homotypic H2A.Z.

Homotypic H2Av nucleosomes are enriched at 5' splice sites

We wondered whether 5'-to-3' polar enrichment of homotypic H2Av nucleosomes is specific to TSSs, or instead represents a general property of nucleosomes undergoing disruption during active transcription. To distinguish between these possibilities, we asked whether similar 5'-to-3' polarity occurs within gene bodies. When we profiled all 67,630 annotated intron/exon junctions we found that homotypic and total H2Av nucleosomes were enriched relative to heterotypic H2Av over the first nucleosome downstream of the 5' intron/exon junction of active genes, decreasing in enrichment within exons for homotypic H2Av (Figure 2-9a-b). Furthermore, the first quintile displayed reduced enrichment for homotypic but not for heterotypic H2Av or for total (input) nucleosomes (Figure 2-10a-c). The enrichment of heterotypic H2Av over exons is most likely accounted for by the higher G+C content of exons relative to introns (Figure 2-10d), and previous work has shown that exons have a higher nucleosome density(Schwartz et al., 2009; Sims et al., 2007; Tilgner et al., 2009). However, the differences between homotypic and heterotypic H2Av must be independent of G+C content, because all five expression quintiles display identical average G+C content over introns and exons, whereas homotypic H2Av displays 5'-to-3' polarity for the first three quintiles relative to the lowest two quintiles. We conclude that the enrichment features of homotypic H2Av nucleosomes are not specific for the boundary between TSSs and gene bodies, but rather are general consequences of encounters between the transcriptional elongation apparatus and regions of high nucleosome density.

Homotypic H2Av nucleosomes are low-salt-soluble

We previously used tiling microarray analysis to show that the distribution of H2Av around TSSs resembles that of low-salt-soluble nucleosomes(Henikoff et al., 2009), where low-

salt solubility defines classical 'active' chromatin(Rocha et al., 1984). We wondered whether this correspondence reflects an inherent property of homotypic H2Av nucleosomes, in which case one might suppose that the different structural properties of H2A.Z versus H2A within the nucleosome core contribute to the generation of active chromatin. To directly compare homotypic H2Av with active chromatin profiles, we prepared low-salt soluble nucleosomes from S2 cells by extraction of intact MNased nuclei with 80mM NaCl and sequenced a library of fragments using the Illumina platform. As previously observed using tiling microarrays, the salt-solubility profile is all-or-none with respect to gene expression, in that the top three quintiles display profiles with peaks of similar height, whereas the lowest two quintiles, which represent inactive genes, display a featureless pattern (Figure 2-9f). The +1 nucleosome downstream of the TSS is enriched in the 80mM soluble fraction, with strong phasing down to the +8 nucleosome and a gradual decay in amplitude. These features of classical active chromatin closely resemble those of both homotypic and total H2Av (Figure 2-9d-e). We then investigated the peak spacing between low-salt-soluble nucleosomes and homotypic H2Av and found that they were virtually identical, both averaging ~175 bp (Table 2-1). We also observed close correspondences between homotypic H2Av, total H2Av and salt-soluble nucleosomes downstream of intron/exon junctions, where successive peaks display similar 5'-to-3' polarity (Figure 2-9c). These results imply that the distinctive physical properties of active chromatin that cause it to be more easily solubilized by low salt arise, at least in part, from structural differences between homotypic H2Av and H2A-containing nucleosomes.

We have previously shown that low-salt-soluble nucleosomes are enriched at epigenetic regulatory elements, including those bound by the trithorax-Group protein, Zeste, and the Polycomb-Group proteins, Enhancer-of-Zeste (EZ) and Posterior-Sex-Combs (PSC). To

ascertain whether a similar enrichment is seen for homotypic H2Av nucleosomes, we aligned sequenced fragments around protein binding sites and averaged over all sites. For the five larger size classes, no enrichment of H2Av of either type or input was seen over Zeste sites and little if any over EZ+PSC sites, although the 55-bp class was characterized by peaks for input, heterotypic, total, and homotypic H2Av nucleosomes (Figure 2-11). This pattern of enrichment at epigenetic regulatory elements is similar to the pattern found at promoters of active genes, where only the 55-bp class is enriched over the TSS (Figure 2-5b). At both epigenetic regulatory elements and TSSs, nucleosomes are relatively depleted, which implies that only particles that are unstable and relatively accessible to MNase are detected, whereas homotypic H2A.Z nucleosomes are especially stable(Thambirajah et al., 2006). Therefore, the enrichment of homotypic H2A.Z nucleosomes does not seem to be a general feature of classical active chromatin, but rather is specific for nucleosome-dense genic regions undergoing transcriptional elongation.

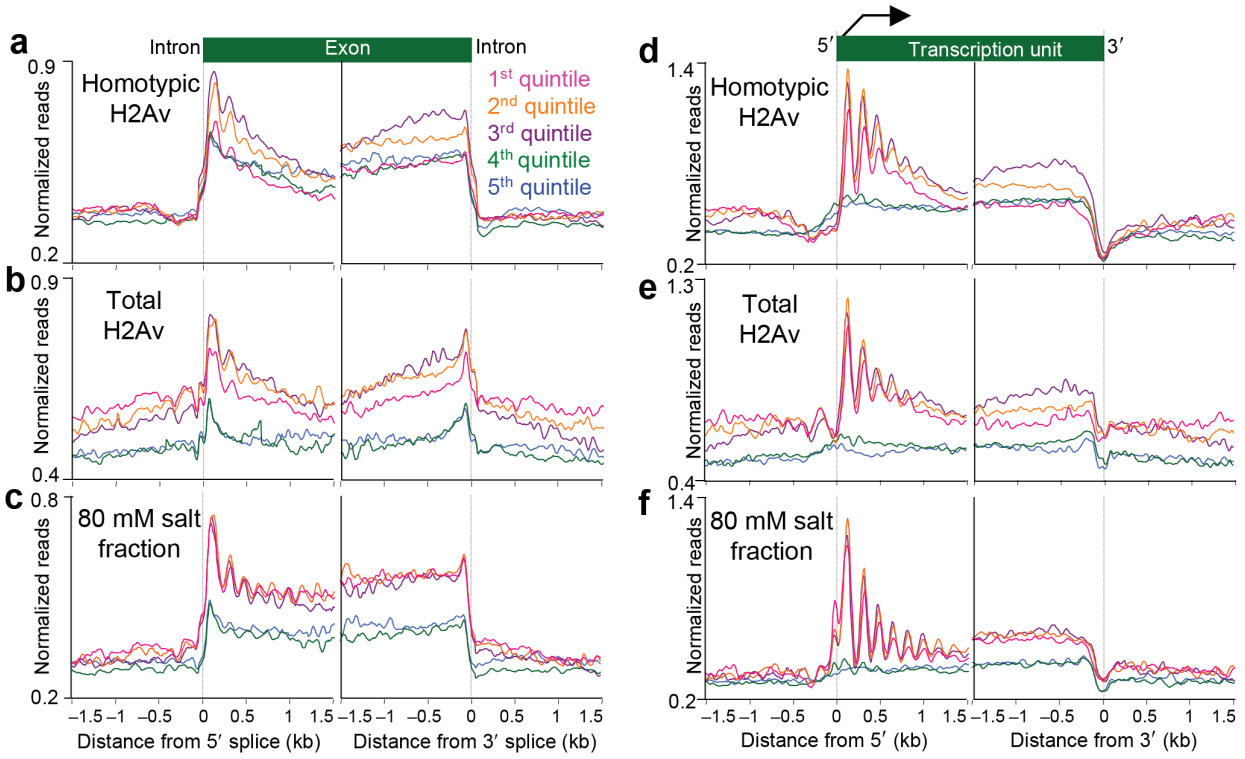


Figure 2-9. Homotypic H2Av nucleosomes are enriched downstream of intron/exon junctions and are low-salt-soluble.
(a) Average profile for homotypic H2Av nucleosomes within the 147-bp size class aligned at intron/exon and exon/intron junctions and split into quintiles of gene expression as in Figure 2-5. (b) Single FLAG-H2Av pulldown (total H2Av) profile for comparison to a. (c) Average profile for DNA from 80mM salt-soluble nucleosomes. Profiles in (a-c) represent means of biological replicates over 67,630 intron/exon junctions, divided into 10-bp bins. (d-f) Same as (a-c) except aligned around gene 5' and 3' ends. The y-axis refers to mapped single-end reads, where the scales were adjusted slightly to facilitate peak comparisons between the panels.

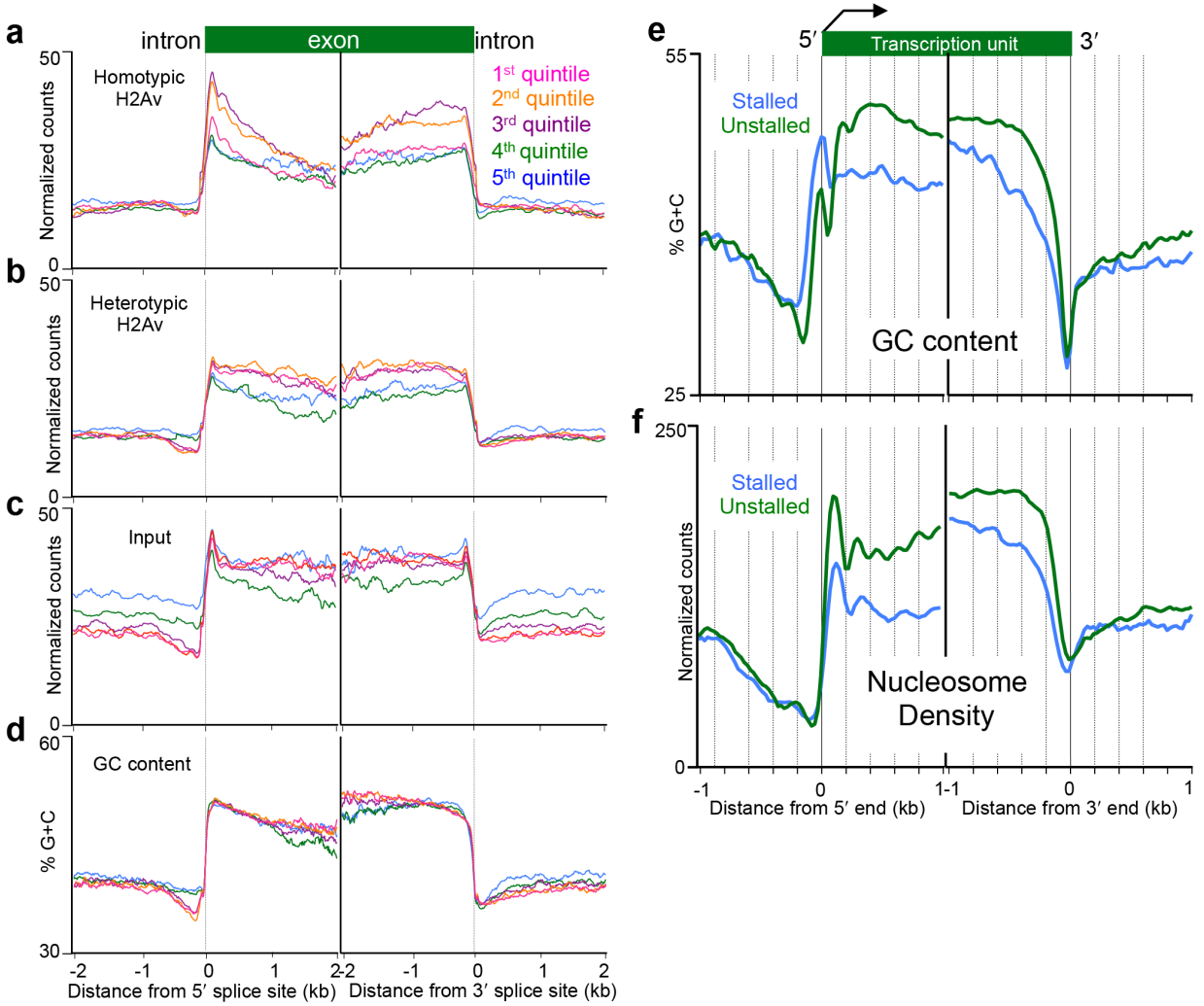


Figure 2-10. Relationship between nucleosome density and G+C content.
Homotypic H2Av nucleosomes are enriched downstream of intron/exon junctions. Average profiles for homotypic (a) and heterotypic (b) H2Av nucleosomes within the 147-bp size class aligned at intron/exon and exon/intron junctions. Genes were split into expression quintiles as in Figure 2-5. Profiles represent means of biological replicates over 67,630 intron/exon junctions, divided into 10-bp bins. (c) Same as b for input nucleosomes within the 147-bp size class. The y-axis refers to paired-end counts. (d) G+C content at intron/exon junctions split into expression quintiles. Averaging was truncated at 5' or 3' ends of neighboring exons. (e-f) Stalled and non-stalled genes differ in G+C content. 950 stalled and 4,061 nonstalled genes were grouped and G+C content determined over successive 10-bp intervals. Nucleosome density is represented by 147-bp counts taken at 25-bp intervals from the mean of paired-end Input reads.

Phased peak distances and spacing

Nucleosome #	Distance from TSS (quintile)			Peak spacing (quintile)		
	1st	2nd	3rd	1st	2nd	3rd
80mM salt fraction						
-1	-170	-170	-170			
PolII?	-20	-20	-20	150	150	150
1	130	130	120	150	150	140
2	310	310	300	180	180	180
3	490	480	470	180	170	170
4	660	650	640	170	170	170
5	830	830	810	170	180	170
6	1010	1000	990	180	170	180
7	1180	1170	1170	170	170	180
8	1350	1340	1340	170	170	170
Mean ±SD (N1->N8)				174.3±5.3	172.9±4.9	174.3±5.3
H2Av Homotypic peaks (Distance and spacing from TSS)						
90-bp interval						
PolII?	-20	-10	-20			
1	150	140	130	170	150	150
2	330	320	290	180	180	160
3	510	490	480	180	170	190
4	680	660	640	170	170	160
Mean (N1->N4)				176.7	173.3	170
130-bp interval						
-1	-170	-170	-170			
1	160	150	140	330	320	310
2	340	330	320	180	180	180
3	510	490	480	170	160	160
4	680	660	650	170	170	170
Mean (N1->N4)				173.3	170	170
147-bp interval						
-1	-170	-170	-170			
1	160	150	140			
2	350	340	330	190	190	190
3	520	500	470	170	160	140
4	690	670	660	170	170	190
Mean (N1->N4)				176.7	173.3	173.3
All size classes (Mean±SD)				175.6±7.3	172.2±9.7	171.1±17.6

Table 2-1

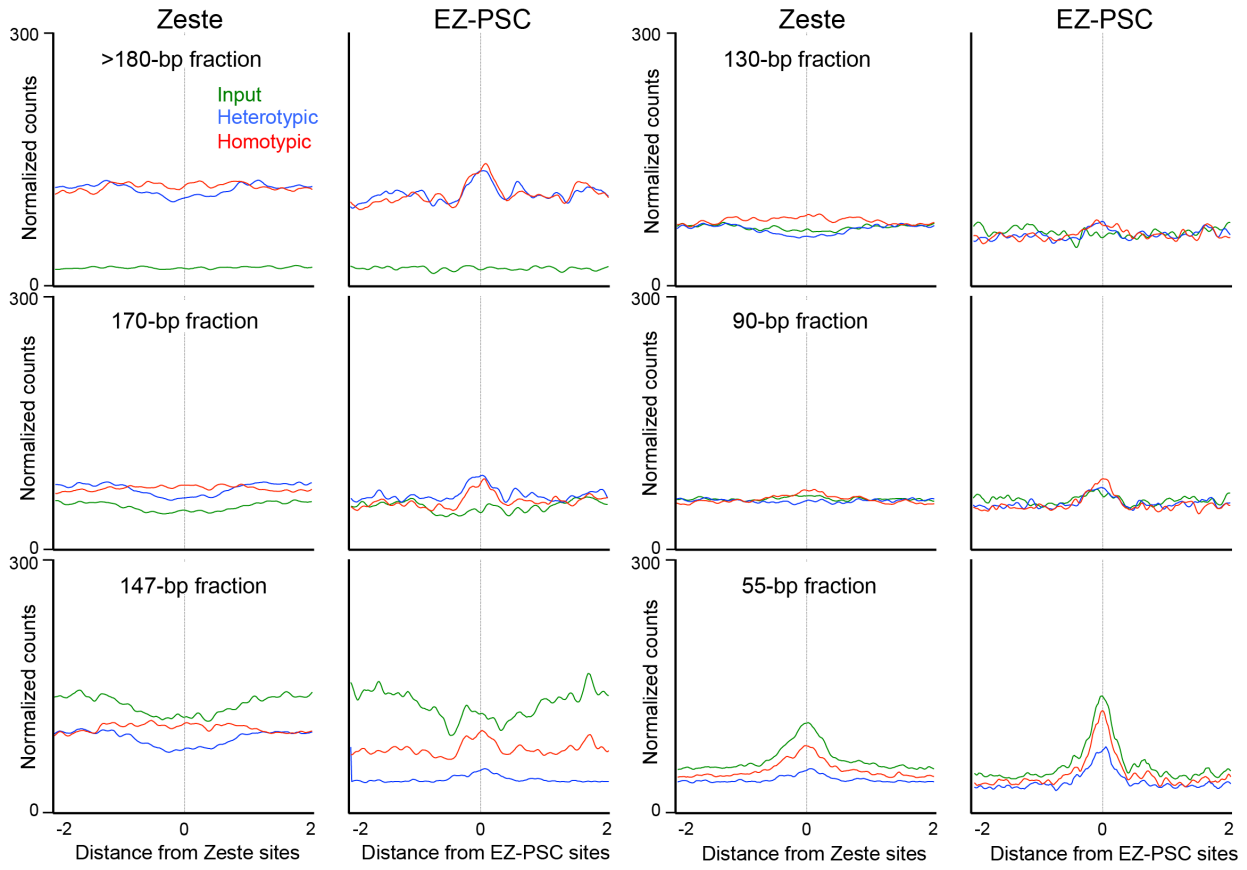


Figure 2-11. Homotypic H2Av is not enriched at epigenetic regulatory elements.
All size classes profiled +/- 1kb from 1428 aligned Zeste (trxG) protein binding sites and 197 Enhancer-of-Zeste and Posterior-Sex-Combs (EZ + PSC = PcG) binding sites for Input, heterotypic and homotypic H2Av nucleosomes at 50 bp resolution.

RNA Polymerase II pauses at H2Av depleted genes

Enrichment of homotypic H2Av nucleosomes where Pol II encounters more densely packed low-salt-soluble nucleosomes raised the question of whether the process of homotypic enrichment is related to the degree of polymerase transit. One situation in which polymerase transit differs between genes is at sites of pausing downstream of the TSS in various animal systems(Muse et al., 2007; Rahl et al., 2010; Rougvie and Lis, 1990). Release of RNA Pol II from the paused state is the rate-limiting step in transcription of these genes. Therefore, we asked whether the genes that are subject to pausing downstream of TSSs in S2 cells are more or less

likely to be enriched in homotypic H2Av than are genes that are not subject to pausing. Using *Drosophila* S2 cell data from a study that identified 1014 stalled genes and 4392 active genes that were not stalled(Muse et al., 2007), we observed a general depletion of nucleosomes over gene bodies in the stalled class relative to those in the unstalled class (Figure 2-12a, bottom panel). We attribute this decrease to the lower G+C content of genes in the stalled class compared to genes in the unstalled class (Figure 2-10a). Relative to this difference, we see a striking enrichment of homotypic and a corresponding depletion of heterotypic H2Av nucleosomes in the unstalled relative to the stalled class (Figure 2-12a, top and middle panels). The lower level of homotypic H2Av in the stalled class relative to the unstalled class is not accounted for by lower overall expression of stalled genes, because this class shows a slight enrichment of expressed genes relative to the unstalled class (Figure 2-13a).

To better understand why homotypic H2Av is depleted from the stalled class, we examined the spacing between homotypic H2Av peaks downstream of the TSSs. It was previously reported that phased H2Av nucleosomes are shifted 10-bp downstream at paused genes compared to those that are not paused(Mavrich et al., 2008). We likewise noted a downstream shift in our homotypic H2Av nucleosome peaks, in that increases in gene expression were correlated with increasing downstream shifts. At the +1 nucleosome, the first expression quintile displayed a 10-bp peak downstream from the second quintile, which displayed a 10-bp shift compared to the third (Figure 2-12b). Similar shifts downstream of the TSS were seen for input nucleosomes (Figure 2-13b), likely reflecting the predominance of homotypic H2Av at these nucleosome positions. Peak shifts increased in length for the +2 and +3 nucleosomes, so that the peaks for the +3 nucleosome was centered 520 bp, 500 bp, and 470 bp downstream of the TSSs for the first, second and third expression quintiles, respectively. These increases in peak

spacing with increasing transcription suggest that homotypic H2Av nucleosomes are displaced downstream during transcriptional elongation. However, no difference in peak shift was detected between nucleosomes in the stalled and unstalled classes (Figure 2-12b and Figure 2-13b). Therefore, the propensity for a homotypic H2Av nucleosome to be displaced downstream during polymerase passage does not seem to depend on whether or not the polymerase is stalled. Rather, this propensity might depend on some inherent property of homotypic H2A.Z nucleosomes.

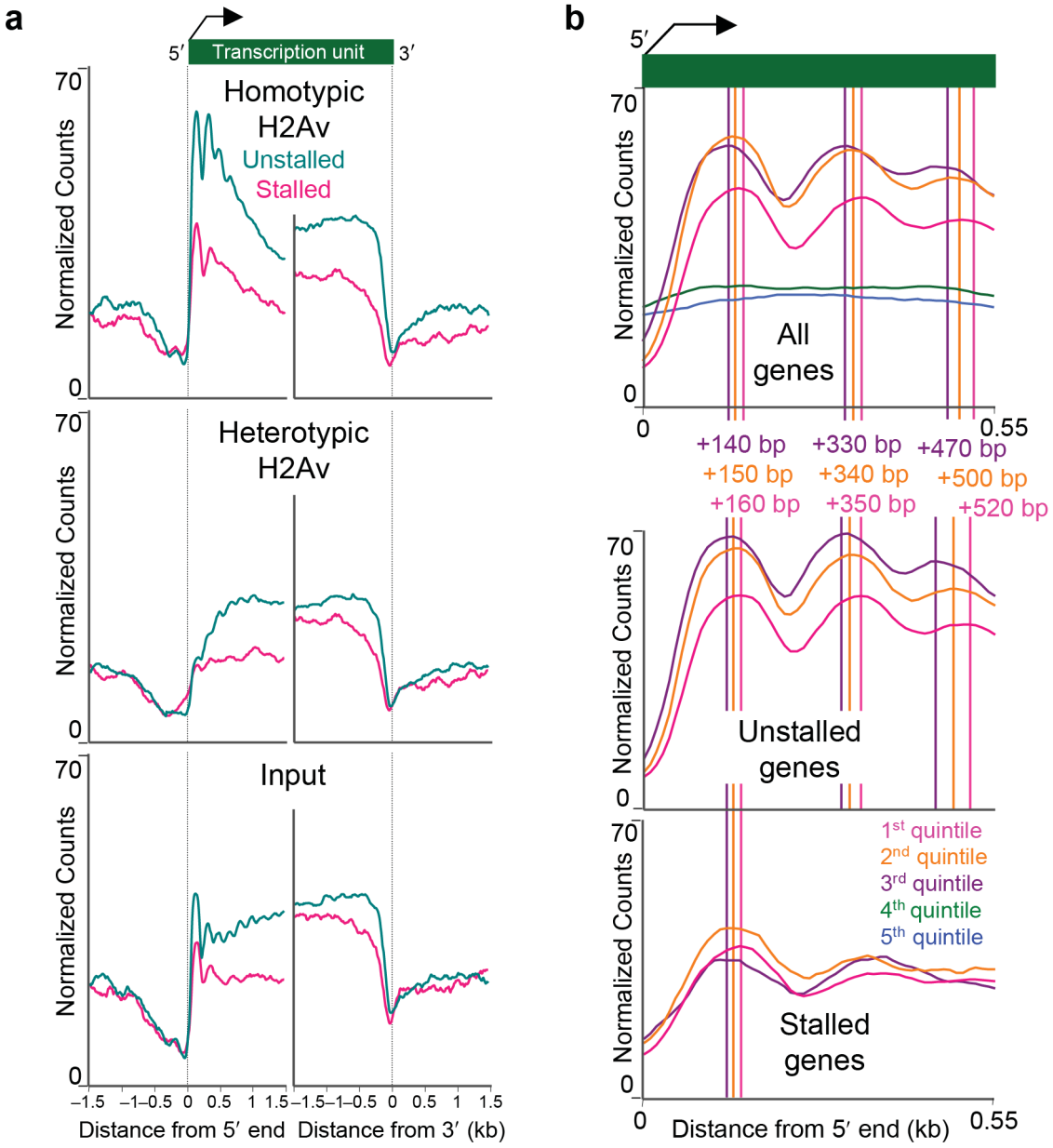


Figure 2-12. Depletion of homotypic H2Av nucleosomes at genes with stalled RNA Polymerase II.

(a) Average profiles for homotypic and heterotypic H2Av and input nucleosomes within the 147-bp size class for 950 genes that have promoter-proximal enrichment of polymerase (stalled, red) and 4061 genes that have polymerase throughout the genes but are not enriched proximal to the promoter (green). (b) Genes with and without stalled Pol II show displacement of homotypic H2Av nucleosomes with increasing expression level. Average profiles for homotypic H2Av nucleosomes within the 147-bp size class show shifting of peaks downstream with increasing expression for the top three quintiles, and relative depletion of the top expression quintile. Similar results are seen for the 950 stalled and the 4061 unstalled genes. Only the +1 position is shown for stalled genes because there are fewer genes and the resolution is lower than for unstalled genes. The y axis refers to paired-end counts.

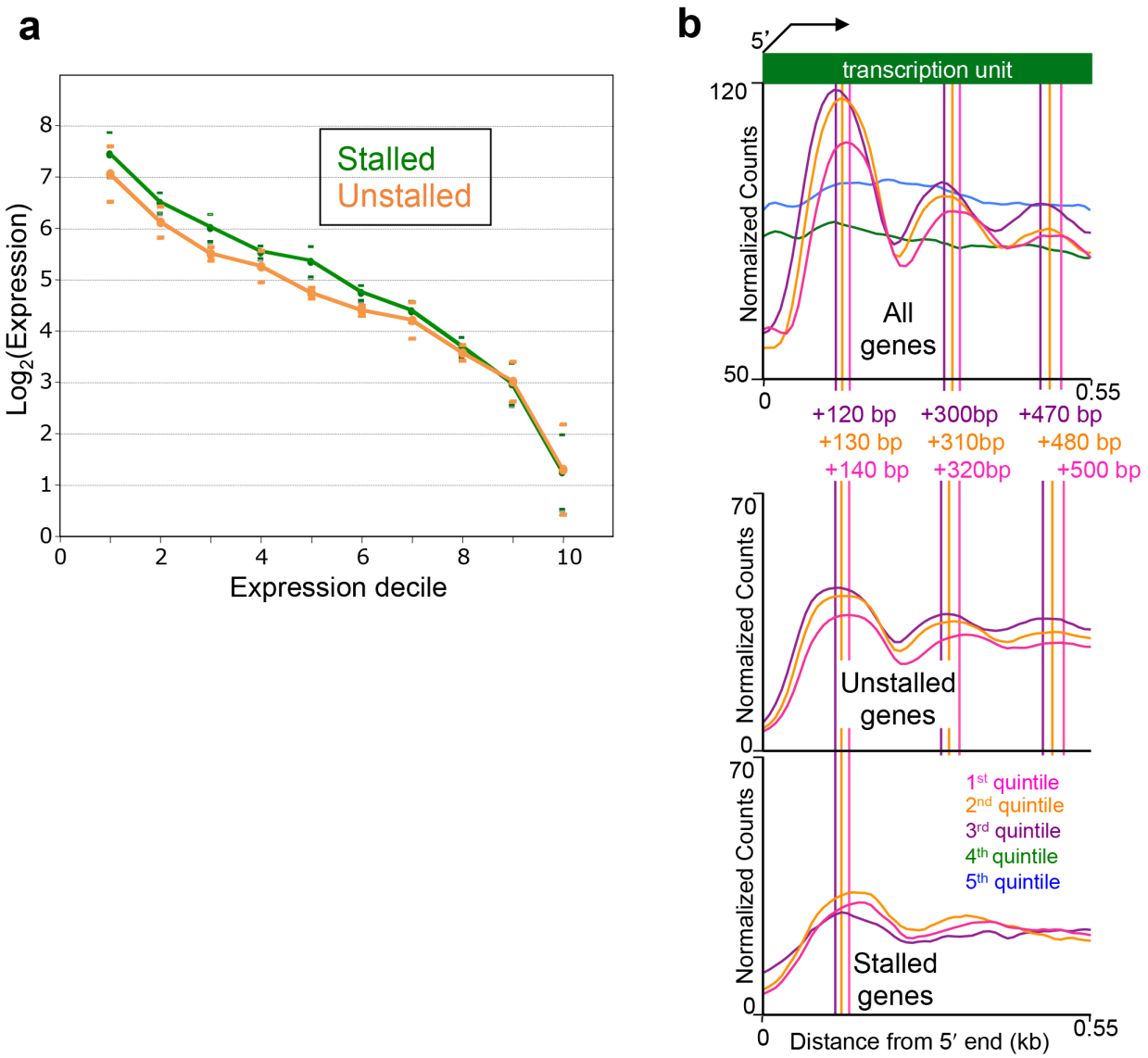


Figure 2-13. Genes with and without stalled Pol II are similarly expressed and show displacement of nucleosomes with increasing expression level.
(a) 950 stalled and 4,061 non-stalled genes were binned into deciles by expression level, averaged, and plotted with corresponding standard deviation. (b) Average profiles for all input nucleosomes show shifting of peaks downstream with increasing expression for the top three quintiles, and relative depletion of the top expression quintile. Similar results are seen for the 950 stalled and the 4061 non-stalled genes. See Figure 2-12 for details.

Active chromatin is depleted from genes with paused RNA Pol II

One property of homotypic H2Av nucleosomes that might contribute to their depletion at paused genes is their low-salt solubility. Indeed, when we examined profiles of low-salt-soluble nucleosomes for stalled and unstalled genes, we detected a striking difference: The stalled gene profile displayed a pair of narrow salt-soluble peaks, at -20 bp and +80 bp relative to the TSS, with very low levels of low-salt-soluble nucleosomes downstream, whereas no such peaks were observed in unstalled genes (Figure 2-14a, dotted blue lines). The 100-bp spacing of these peaks is less than the 150-bp spacing that we would expect for bona fide nucleosomes, which we had assumed in combining plus- and minus-strand reads to display the 80mM single-read data (Figures 2-9a and 2-9d). Therefore, we asked whether these peaks instead represent protection by a particle located directly in between the two peaks, centered over +30 but only in genes with paused Pol II. To evaluate this possibility, we mapped the ends of the plus and minus reads without any offset to compensate for peak spacing. This revealed that their distributions approximately coincide around a peak centered over +30 (Figure 2-14, top panels, green and orange lines), rather than showing further peak splitting as would be expect for phased nucleosomes. The coincidence of plus- and minus-strand reads without offset implies that they result from protection of DNA from MNase digestion by a very small particle that is located on average at +30.

To confirm that the the protected particle is at the precise location of paused Pol II, we performed limited paired-end sequencing of salt-soluble chromatin. By mapping and binning reads in the 25-75 bp range, we detected a sharp peak ~50-bp wide centered over +30 for active genes in the stalled class, with a relatively small peak at this location for genes in the unstalled class (Figure2-14b, bottom panels). As +30 is the mean location of paused RNA Pol II(Muse et

al., 2007), we infer that our sequence mapping of classical active chromatin has accurately footprinted this species, which implies that paused Pol II is itself enriched in low-salt-soluble chromatin(Henikoff et al., 2009).

We also observed that phased low-salt-soluble nucleosomes downstream of paused Pol II were strongly depleted relative to those in the unstalled class (Figure 2-14b). This implies that pausing inhibits the generation of salt-soluble active chromatin downstream. Therefore, genes with paused Pol II are deficient for both homotypic H2Av nucleosomes and low-salt-soluble nucleosomes, which provides additional evidence that homotypic H2A.Z is a key attribute of transcriptionally active chromatin.

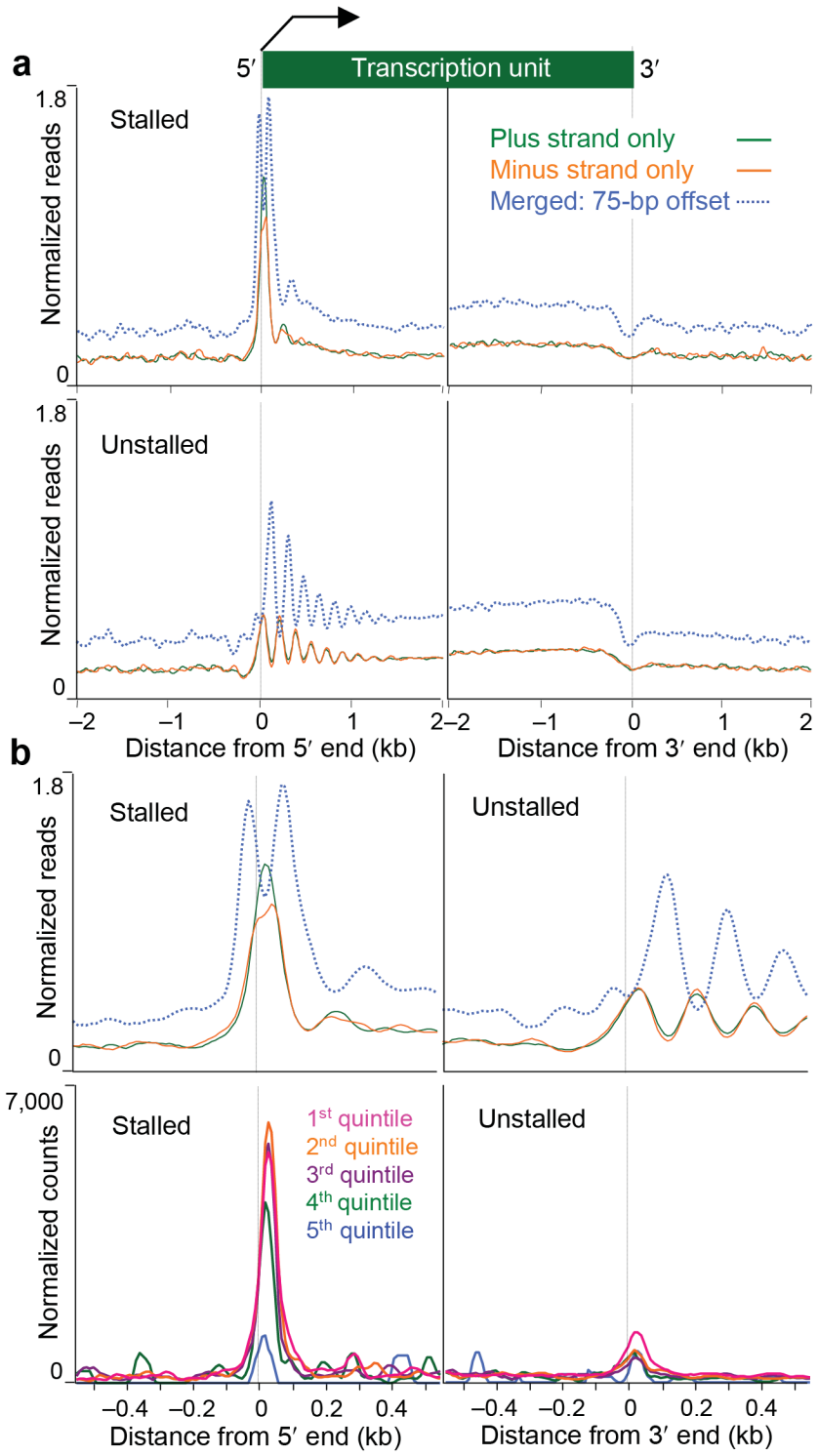


Figure 2-14. Low-salt extracted chromatin reveals distinctive features of RNA Pol II stalling.

Single-end sequence reads from DNA extracted from low-salt-soluble nucleosomes were displayed for 950 stalled and 4061 unstalled genes (a). The mapped ends of plus-strand reads were offset by adding 75 bp, and minus-strand reads were offset by subtracting 75 bp (dotted blue lines), which assumes that all reads represent one end or the other of a protected

nucleosome core. Read ends were mapped separately for plus strands (green lines) and minus strands (orange lines) without offset, which assumes that all reads were derived from a small protected fragment. Average profiles are shown for low-salt-soluble nucleosomes mapped over genes in the stalled and unstalled classes. The y-axis refers to mapped single-end reads. As a negative control, we also performed the same analysis at intron/exon junctions, but did not observe evidence of enhanced protection at the corresponding position in comparing stalled and unstalled profiles (Figure 2-15). (b) A biological replicate sample of low-salt extracted chromatin was subjected to limited paired-end sequencing. The 35–55-bp fraction is displayed in the bottom panels for quintiles of expression for genes in the stalled and unstalled classes. For comparison, the 5' plots from panel (a) is shown in the panels above on the same x-axis scale. The y-axis refers to paired-end counts.

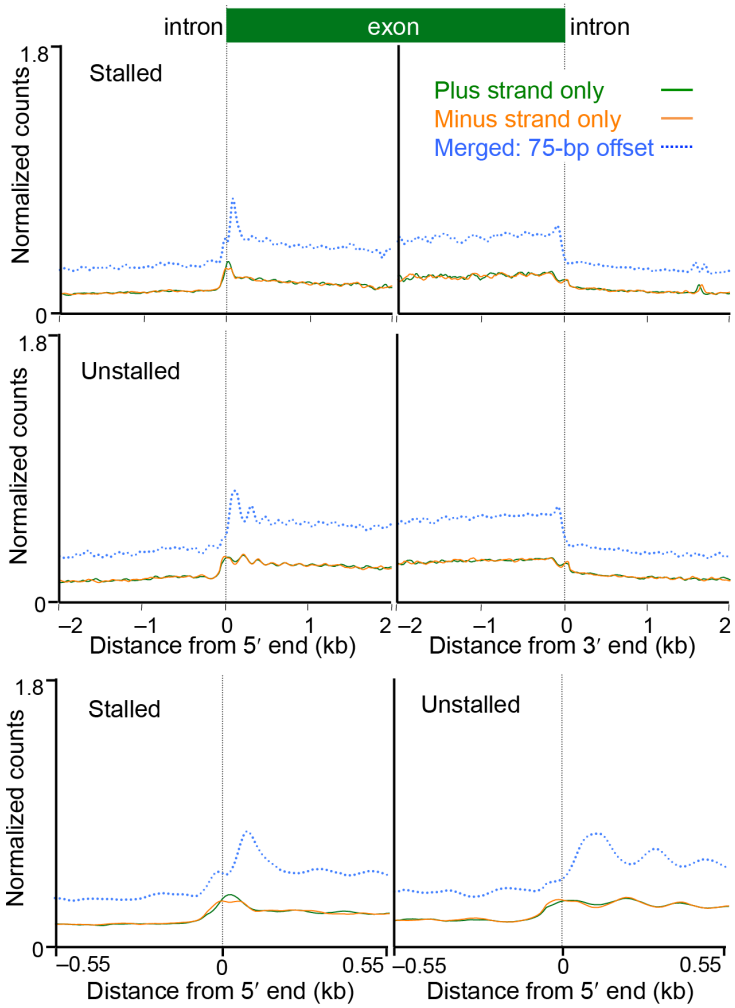


Figure 2-15. Short low-salt soluble fragments are not found at intron-exon junctions. Similar analysis as shown in Figure 2-14 but at intron/exon boundaries. Note that there is little difference between stalled and non-stalled landscapes, in contrast to the situation at +30, where only genes with stalled Pol II show protection (Figure 2-12).

Discussion

Our mapping of individual H2Av nucleosomes across the *Drosophila* genome has revealed unexpected differences between homotypic and heterotypic H2A.Z nucleosomes. First, only homotypic H2A.Z nucleosomes are enriched over genes, whereas heterotypic nucleosomes are depleted. Second, a physical property of chromatin, namely its solubility, co-localizes with the distribution of homotypic H2A.Z. Third, both homotypic H2A.Z and low-salt solubility of nucleosomes correspond to one another at intron/exon junctions in the same manner as at transcriptional start sites, which implies that properties of homotypic H2A.Z nucleosomes result from RNA Polymerase II transit. Fourth, genes with paused Pol II are depleted in both homotypic H2A.Z and low-salt-soluble nucleosomes, further suggesting that homotypic H2A.Z nucleosomes and classical active chromatin are generated during transcriptional elongation. These findings have implications both for understanding the mechanism whereby H2A.Z becomes preferentially enriched at active genes and for how H2A.Z might function in transcriptional regulation.

Homotypic H2A.Z nucleosomes differ structurally from canonical H2A nucleosomes(Suto et al., 2000), and we suggest that these differences in localization and *in vivo* properties result from an underlying structural difference between nucleosomes of different composition. An obvious candidate for such a structural difference is the interface between H2A/H2B dimers within the nucleosome core(Henikoff, 2009; Thakar et al., 2010). This dimerization interface is more extensive for H2A.Z/H2A.Z than for H2A/H2A, and there is a potential structural clash between H2A.Z and H2A in the same core particle(Suto et al., 2000). However, heterotypic H2A.Z nucleosomes form readily *in vitro*(Chakravarthy et al., 2004) and are abundant *in vivo*(Viens et al., 2006), and we have found that they are broadly distributed

genome-wide. Alternatively, the conserved differences in the docking domains of H2A.Z and H2A might be responsible for their differences in distribution and *in vivo* properties. In either case, heterotypic H2Av nucleosomes might be sufficiently stable to assemble, but more likely than homotypic H2Av nucleosomes to unravel when disrupted. Nucleosomes are disrupted over promoters of active genes, where unstable core particles containing both H3.3 and H2A.Z from human cells have been found to be enriched(Jin et al., 2009). It is possible that these unstable double-variant nucleosomes are more accessible to internal cleavage by MNase, and so might correspond to the 55-bp class of TSS-derived small DNA fragments that we obtained from H2Av-containing nucleosomes (Figure 2-9b, bottom panels).

The enrichment of H2A.Z downstream of TSSs and intron/exon junctions might be understood by a simple model based on relative stability of homotypic H2A.Z nucleosomes to disruption by RNA Pol II transit, and the *in vitro* observation that RNA Pol II evicts an H2A/H2B dimer when it transcribes through a nucleosome(Kireeva et al., 2002) (Figure 2-16). When a dimer is occasionally lost at random *in vivo*, the resulting hexameric particle becomes a substrate for new dimer addition, which can be either H2A-H2B or H2A.Z-H2B. If replacement is by the Swr1 complex then an H2A.Z-H2B dimer will be deposited(Mizuguchi et al., 2004), forming a heterotypic nucleosome. A subsequent round of transcriptional elongation would lead to occasional loss of either dimer. When the loss is of the remaining H2A-H2B followed by the gain of a second H2A.Z-H2B, a homotypic H2A.Z nucleosome will result. One possibility is that the more extensive dimerization interface between the two H2A.Zs in the core particle(Suto et al., 2000) would make them more stable than either canonical H2A/H2A or heterotypic H2A.Z/H2A to unraveling and losing a dimer by subsequent Pol II encounters. Alternatively, the docking domain of H2A.Z might interact more stably with H3 and thus show a lower propensity

to be lost than H2A. Thus, the more rounds of transcription through a nucleosome position, the more likely it will be that a nucleosome will be homotypic for H2A.Z and the more likely that it will be extracted by low-salt because it is repeatedly disrupted. In this way the unique intranucleosomal interaction between H2A variants would ultimately be responsible for altering the physical properties of chromatin.

Our observation that homotypic H2A.Z nucleosomes are depleted over genes with stalled polymerases argues against proposals that H2A.Z nucleosomes act as simple barriers to transcriptional elongation(Mavrich et al., 2008; Thakar et al., 2010). Rather, our results suggest that homotypic H2A.Z nucleosomes facilitate, not impede, transcriptional elongation. They might do this, for example, by recruiting ATP-dependent nucleosome remodelers, such as ISWI, which preferentially mobilizes H2A.Z nucleosomes(Goldman et al., 2010). We speculate that when homotypic H2A.Z nucleosomes are mobilized, they facilitate polymerase transit by helping to relieve torsional strain. RNA polymerase creates positive torsion ahead and negative torsion behind as it moves a denatured DNA bubble forward. When it approaches a nucleosome, the positive torsional stress helps to unravel the left-handed nucleosome core particle(van Holde et al., 1992). At the same time, the negative torsional stress in the wake of polymerase helps to rewind the DNA around the nucleosome core. A wave of positive torsional stress ahead of and negative torsional stress behind elongating RNA Pol II is thought to facilitate nucleosome disassembly and reassembly(Bancaud et al., 2007; Kouzine et al., 2008; Wang and Droge, 1996). We propose that the distinct interaction surfaces of H2A.Z facilitate nucleosome reformation in the wake of elongating polymerases, thus more rapidly taking up negative supercoils and relieving the negative strain caused by polymerase transit. Relief of negative strain is expected to reduce backtracking of RNA Polymerase II, which is thought to be the mechanistic basis for

polymerase stalling(Galburt et al., 2007). Thus, enrichment of homotypic H2A.Z nucleosomes at TSSs would be a general adaptation for relieving paused polymerase at TSSs and downstream.

The enrichment of Drosophila H2Av over phased nucleosomes is typical for H2A.Z patterns seen in plants, fungi and other animals(Albert et al., 2007; Barski et al., 2007; Lantermann et al., 2010; Zilberman et al., 2008). However, *Drosophila* is unique in that its H2A.Z has acquired a functional C-terminal SQ(E/D)F motif (where F represents a hydrophobic residue), which in other eukaryotes is present on H2A.X, a replication-independent variant that belongs to the H2A clade(Talbert and Henikoff, 2010). The function of the SQ(E/D)F motif is well-established: SQ(E/D)F serines in the vicinity of a double-strand (ds) break become rapidly phosphorylated, and the resulting g–H2AX phosphates help to recruit the homologous ds DNA repair machinery to facilitate accurate repair(Fernandez-Capetillo et al., 2004). Because ds breaks can occur anywhere, H2A.X must be ubiquitously distributed throughout the genome. Nucleosome disruption during replication fork passage might result in such a ubiquitous H2Av distribution genome-wide. Thus, the uniformity of heterotypic H2Av throughout the genome might be seen as a feature that has allowed *Drosophila* to dispense with its ancestral H2A.X after acquisition of the SQ(E/D)F motif at the C-terminus of its only other replication-independent H2A variant. Our results imply that only homotypic H2Av participates in transcriptional regulation, whereas the relative uniformity of heterotypic H2Av may have permitted the acquisition of a ubiquitous DNA repair function. Thus, the distinct structural features of heterotypic and homotypic H2Av might have led to the evolution of H2Av bifunctionality.

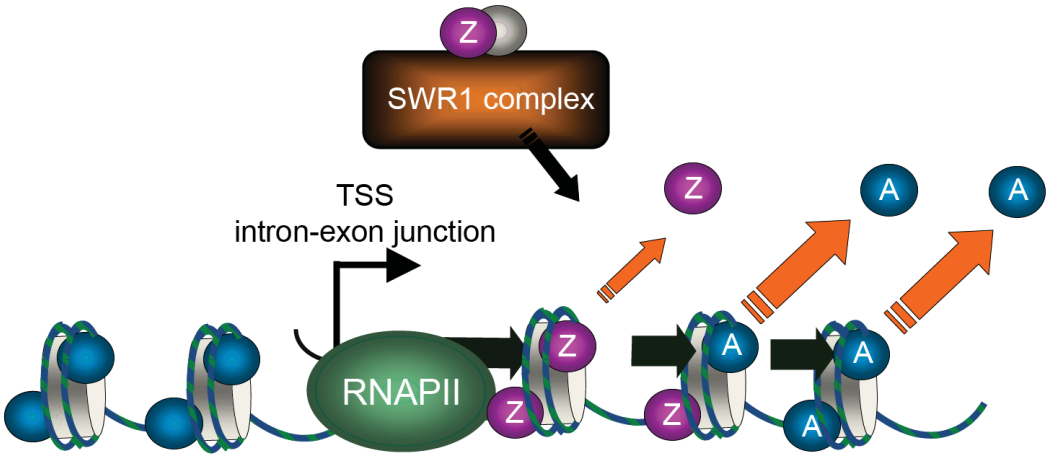


Figure 2-16. Model for the generation of H2A.Z enrichment patterns.
The cartoon depicts nucleosome disruptions that occur upon RNA Polymerase II transit. Pol II disruption causes dimer loss that is less for homotypic H2A.Z than for either heterotypic, or canonical H2A, with homotypic H2A.Z being the most stable nucleosome to disruption. H2A.Z homotypics are formed with multiple rounds of Pol II transit and remain relatively stable to further disruption. Heterotypics form but are relatively unstable to successive rounds of Pol II transit, leading to their depletion relative to homotypics. H2A.Z recruits remodelers which should exhibit the highest activity on homotypic H2A.Z, in support of transcription.

Methods

Cell Culture and Transgene Expression.

Stable Drosophila S2 cell lines were created as described(Henikoff et al., 2009; Mito et al., 2005). We expressed FLAG-tagged H2Av in these lines with baculovirus under the metallothionein promoter as described(Bryson et al., 2010). Briefly, we infected one 150cm2 flask of 6 x107 BLRP-H2A and BLRP-H2Av cells and one flask of S2 cells with a baculovirus construct encoding FLAG-H2Av for two hours at a multiplicity of infection of 100. We then added fresh media and induced transgene expression with 0.5mM CuS04 for 72 hours.

Chromatin Preparation.

We prepared chromatin for affinity purification essentially as described(Bryson et al., 2010), except that we fixed soluble nucleosomes with formaldehyde prior to affinity purification. Briefly, ~3x108 cells from each flask were suspended in HM2 (10mM Hepes, 2mM MgCl2, 0.5mM phenylmethanesulphonylfluoride (PMSF)) and lysed with 0.6% (v/v) NP-40. Nuclei were resuspended in 900µl HM2 and CaCl2 was added to ~1mM, and warmed to 37° for 2'. We digested chromatin with 1U of MNase for 10', and stopped the reaction by adding 5mM EDTA and incubated for 30' on ice. We checked for nuclear retention of total chromatin on protein gels, and detected only ~1.5% of the total released into the soluble fraction (Supplementary Fig. 11). Nuclei were then resuspended in HE buffer (10mM Hepes 0.25mM EDTA 0.5mM PMSF) and chromatin was needle extracted(Jin and Felsenfeld, 2007). Soluble chromatin following MNase and needle extraction was diluted to an OD260 of 3-3.5 (10mm path) with HE, fixed for 1hr at RT with 1% (w/v) fresh formaldehyde (pH 7.0), quenched with 125mM glycine, and dialyzed for 1hr with 3 changes of HE buffer at 4°. Overall recovery of MNased nucleosomes averaged ~60%. We removed 150µl to serve as input chromatin and used the remainder for affinity purification. We found that tagged H2Av nucleosomes included the same approximate ratio of tagged and untagged H2A as input chromatin (Supplementary Fig. 1c).

We prepared 80mM soluble chromatin as described(Henikoff et al., 2009). We found that profiling of H2Av-containing nucleosomes from this material led to highly inconsistent results (S. H., unpublished results), perhaps owing to the reported instability of double-variant H3.3+H2A.Z nucleosomes when subjected to such salt treatments(Jin et al., 2009).

Chromatin affinity purification.

We incubated chromatin with 50µl of equilibrated Anti-FLAG affinity agarose beads (Sigma, A2222) overnight at 4°C with end-over-end rotation. Beads were collected by centrifugation, the unbound fraction was collected, and the beads were washed with HE supplemented with 150mM NaCl and 0.1% Triton X-100 (v/v), 3x and then 2x with HE buffer, 10' on a rotator for each wash to yield single-FLAG-H2Av pulldown material (total H2Av). For sequential affinity purification we eluted the bound FLAG-H2Av nucleosomes with 200µl of 100µg ml-1 3xFLAG peptide in HE, 30' at 4°, the supernatant was collected, and the elution was repeated 2 x 10' at RT. The beads were then washed with 400µl HE and the supernatant was added to eluate. We then centrifuged the entire eluate 5' at 5,000 RPM, and the supernatant was incubated with 50µl packed streptavidin-sepharose (GE) for 3hr on a rotator at 4°C and washed as before.

Illumina library construction and sequencing.

Affinity purified nucleosomes from a single FLAG-H2Av (total H2Av) and two biological replicates for input, heterotypic, and homotypic purification were incubated with 0.1M NaCl and 10mM EDTA for 4 hours at 65° to reverse the formaldehyde crosslinks. DNA was purified as described(Henikoff et al., 2009), except that the Proteinase K digestion was extended to 1 hour. We prepared libraries with 100ng of DNA as described (Illumina) and size-selected our libraries after linker addition to a total length of between 100-600 bp. Samples were sequenced in paired-end mode (36-bp read length). For the 80mM soluble chromatin we prepared libraries from 500ng of DNA as described (Illumina) and size-selected our libraries to between 100-600 bp. This library was sequenced in single-end mode (35-bp read length).

Data Processing.

For paired-end data we aligned uniquely mapped 35-bp reads with paired-end alignment score greater than zero to Release 5 of the Drosophila melanogaster reference sequence (http://flybase.org) using Eland (Illumina, Inc.). We divided each interval count by the total number of counts and rescaled by multiplying by one billion. We averaged the two input, heterotypic and homotypic H2Av biological replicates.

To analyze reads by length, we split up our paired-end fragments into the following size classes: 55-bp (35–75bp), 90-bp (76–110bp), 130-bp (111–140bp), 147-bp (141–160bp), 170-bp (161–180bp), > 180-bp (181–333bp) and repeated the above processing for each size class, normalizing to the total number of counts for the size class.

For single-end data, we used all uniquely mapped reads and selected one location at random for reads that mapped to more than one location on the genome. We moved the position of each read to the middle of a 150-bp fragment, fit a Gaussian kernel density function with bandwidth 20 bp at 10-bp intervals for each chromosome, and normalized so that the values for each 10-bp interval added up to the number of reads for the chromosome. For individual strand analysis, we used only the reads mapped to the plus or minus strand and did not re-locate the read position before fitting the kernel density function(Gehring et al., 2009).

Data Analysis.

To analyze the ends of genes, we took 10,997 genes with both the 5’ and 3’ ends that are annotated in FlyBase r5.23. If genes had more than one transcript we used the longest one. We then rank-ordered genes based on expression level in S2 cells(Henikoff et al., 2009), split these genes into quintiles, and calculated the average number of normalized reads using 10-bp bins. Contributions from neighboring transcription units were omitted during averaging. For stalled

and unstalled genes we matched 950 stalled and 4,061 unstalled genes(Muse et al., 2007). We repeated the ends analysis but filtered the gene to these two sets. We profiled 67,630 unique exons annotated in FlyBase r5.23 using chromosomes 2L, 2R, 3L, 3R, 4, and X only, similarly to ends. Transposons were analyzed essentially as previously described(Henikoff et al., 2009). For analysis of epigenetic regulatory elements, we used published Zeste(Moses et al., 2006) and EZ-PSC(Schwartz et al., 2006) ChIP-chip binding data as previously described(Henikoff et al., 2009), and averaged the paired-end data for all size classes at 50-bp resolution.

Accession number

Genomic Data presented in this work are available from Gene Expression Omnibus under accession number GSE21615.

Chapter 3

Nucleosomes are context-specific, H2A.Z modulated barriers to RNA polymerase

Modified from a paper published in Molecular Cell

Abstract

Nucleosomes are barriers to transcription *in vitro*, however, their effects on RNA polymerase *in vivo* are unknown. Here we describe a simple and general strategy to comprehensively map the positions of elongating and arrested RNA Polymerase II (RNAPII) at nucleotide resolution. We find that the entry site of the first (+1) nucleosome is a barrier to RNAPII for essentially all genes, including those undergoing regulated pausing farther upstream. In contrast to the +1 nucleosome, gene body nucleosomes are low barriers and cause RNAPII stalling both at the entry site and near the dyad axis. The extent of the +1 nucleosome barrier correlates with nucleosome occupancy but anti-correlates with enrichment of histone variant H2A.Z. Importantly, depletion of H2A.Z from a nucleosome position results in a higher barrier to RNAPII. Our results suggest that nucleosomes present significant, context-specific barriers to RNAPII *in vivo* that can be tuned by the incorporation of H2A.Z.

Introduction

Eukaryotic transcription by RNA polymerase II (RNAPII) is highly regulated and occurs on a DNA template that is condensed into repeating nucleosome units, each consisting of 147 bp of DNA wrapped around an octamer of histone proteins. Nucleosomes regulate access of cellular machinery to DNA and are modified by incorporation of functionally distinct histone variants (Talbert and Henikoff, 2010). Transcription across nucleosomes has been extensively

studied *in vitro* (Kulaeva et al., 2013). In this defined setting, nucleosomes present a physical barrier that causes backtracking/arrest of RNAPII and this barrier cannot be efficiently overcome unless the nucleosome is destabilized either with high salt or ionic detergent. These studies typically use minimal components to reconstruct transcription, whereas transcription *in vivo* is regulated by many protein complexes that modify the nucleosomal barrier. In this more complex setting, the effect of canonical and variant nucleosomes on RNAPII elongation is unknown.

The primary approach to investigate transcription in vivo has been to map the position of RNAPII on genomic DNA. However, progress in the field has been limited by methodologies, such as chromatin immunoprecipitation (ChIP), that have insufficient resolution. Recently, single nucleotide resolution strategies for mapping RNAPII were introduced, but these have specific limitations for investigating the role of nucleosomes in transcription. Native elongating transcript sequencing (NET-seq) (Churchman and Weissman, 2011) maps both elongating and backtracked/arrested complexes. However, NET-seq requires solubilization of RNAPII complexes, which is typically far from complete under native conditions in metazoan cells (Kimura et al., 1999). Another strategy, Precision Nuclear Run-on sequencing (PRO-seq) maps elongation competent RNAPII, but cannot detect backtracked/arrested complexes and involves removal of chromatin proteins (Kwak et al., 2013).

In budding yeast, NET-seq detected barriers to RNAPII elongation over gene body nucleosomes but these did not precisely match the consensus locations found *in vitro* (Bondarenko et al., 2006; Jin et al., 2010). Importantly, the transcription start site (TSS) overlaps the first nucleosome (+1) in budding yeast, preventing analysis of RNAPII elongation across this position (Rhee and Pugh, 2012). However, the +1 nucleosome might play an important role in transcriptional regulation (Cairns, 2009) and in most organisms, the +1 nucleosome is found

downstream of the TSS, which is relatively depleted of nucleosomes (Jiang and Pugh, 2009). The +1 nucleosome is a strong barrier to transcription *in vitro* (Nock et al., 2012). This barrier is partially relieved by TFIIS mediated reactivation, implying that RNAPII backtracks after encountering the nucleosome. Using PRO-seq to map elongating Drosophila RNAPII *in vivo*, the effect of the +1 nucleosome or downstream nucleosomes on transcription was inconclusive (Kwak et al., 2013). Thus, it remains unclear whether the nucleosome barrier observed *in vitro* exists *in vivo*, mainly due to technical limitations of current techniques to map RNAPII.

In all eukaryotes, the +1 nucleosome and a few downstream nucleosomes of active genes are enriched for the histone variant H2A.Z (Bonisch and Hake, 2012). H2A.Z is incorporated outside of replication and is mostly essential for viability across eukaryotes (Zlatanova and Thakar, 2008). In budding yeast, H2A.Z deletion led to slower RNAPII elongation at a single fusion gene (Santisteban et al., 2011) while *in vitro*, human H2A.Z nucleosomes were completely refractory to transcription (Thakar et al., 2010). Thus, the effect of H2A.Z on transcription across nucleosomes remains unclear.

Here we examine how RNAPII transcribes through nucleosomes *in vivo* and the effect of H2A.Z incorporation on RNAPII elongation. We describe a simple, generalizable strategy that comprehensively maps all forms of chromatin-bound RNAPII at nucleotide resolution without requiring transgenes, solubilization, or immunopurification. Using this strategy, we show that nucleosomes form context-specific barriers to transcription that can be tuned at least in part by incorporation of H2A.Z and provide insight into how RNAPII transcribes through nucleosomes *in vivo*.

Results

Comprehensive capture and precise strand-specific mapping of RNAPII

To comprehensively and precisely map total RNAPII, we took advantage of the extraordinary stability of the RNAPII ternary complex, which once formed, remains bound to DNA in the presence of high salt, urea, detergents, and polyanions (Cai and Luse, 1987; Wuarin and Schibler, 1994). In metazoans, the complex is also highly insoluble even after nuclease digestion of genomic DNA (Kimura et al., 1999), precluding the use of NET-seq, which requires complete solubilization for RNAPII ChIP. Instead, we utilized the insolubility of RNAPII to purify the engaged complex and its attached nascent chain away from soluble RNA (Figure 3-1A). We quickly arrested transcription in *Drosophila* S2 cells with cold and used buffers containing nonionic detergents and EDTA to sequentially lyse the cell and the nucleus (Mendez and Stillman, 2000; Nishino et al., 2012; Wysocka et al., 2001), leaving an insoluble chromatin pellet with bound RNAPII and associated nascent RNA. Even though the ternary complex is resistant to harsh conditions, we sought to minimize bias associated with the release of unstable complexes by using minimally disruptive conditions. Using this protocol we successfully obtained comprehensive RNAPII recovery in the insoluble chromatin pellet (Figure 3-1B). We isolated nascent RNA from the insoluble pellet and selected for the presence of a 5' 7-methylguanosine cap to further enrich for mRNA transcripts (Nechaev et al., 2010) and sequenced the 3' end to determine the precise position of RNAPII using an established protocol (Churchman and Weissman, 2012) (Figure 3-1C). This approach provides strand-specific, reproducible maps of the active site of RNAPII at single nucleotide resolution (Figure 3-2A, see also Figure 3-3 and Table 3-1). A high correlation between the mapped strand of the 3' end of Nascent Transcripts (3'NT) and the annotated gene model strand, in addition to a relatively uniform distribution of 3'NT signal between introns and exons, indicates that we are mapping

nascent transcripts. Additionally, we observed a good correlation between the 3'NT read density and crosslinked ChIP-seq enrichment of RNAPII (r=0.69) (Figure 3-2B). We also observed large 3'NT peaks at single positions, indicative of RNAPII stalling at individual positions during transcription. Thus, by this simple and general approach, we can comprehensively map RNAPII at nucleotide resolution, enabling investigation into the effect of nucleosomes on transcription.

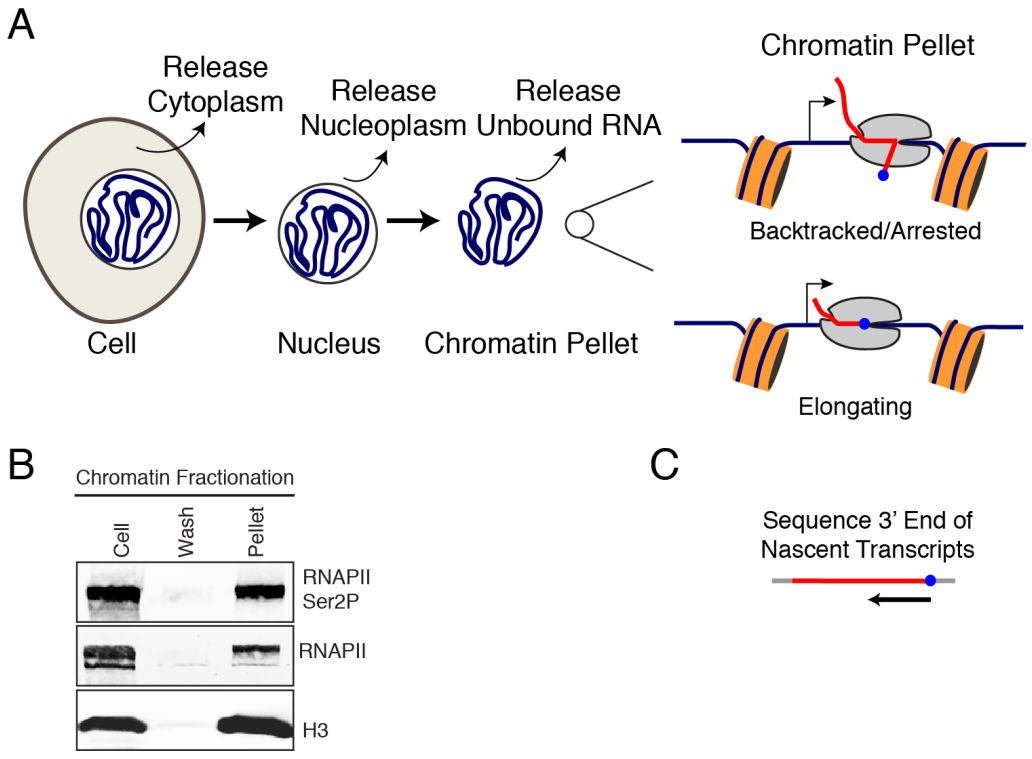


Figure 3-1. Approach to comprehensively capture RNAPII and nascent RNA
(A) A gentle cellular fractionation scheme captures total bound RNAPII with nascent RNA, which is then sequenced to determine the precise 3' end of nascent transcripts (blue dot). The protocol uses hypotonic buffers and non-ionic detergents to release cytoplasm, nucleoplasm and unbound RNA, capturing bound RNAPII complexes and nascent RNA in the insoluble chromatin pellet. (B) A representative immunoblot of Ser2-phosphorylated RNAPII, RNAPII CTD, and H3 load control from whole cell, cumulative wash, and final insoluble pellet fractions, showing high RNAPII recovery in the pellet. (C) Nascent transcripts are extracted and their 3' ends are determined by Illumina sequencing.

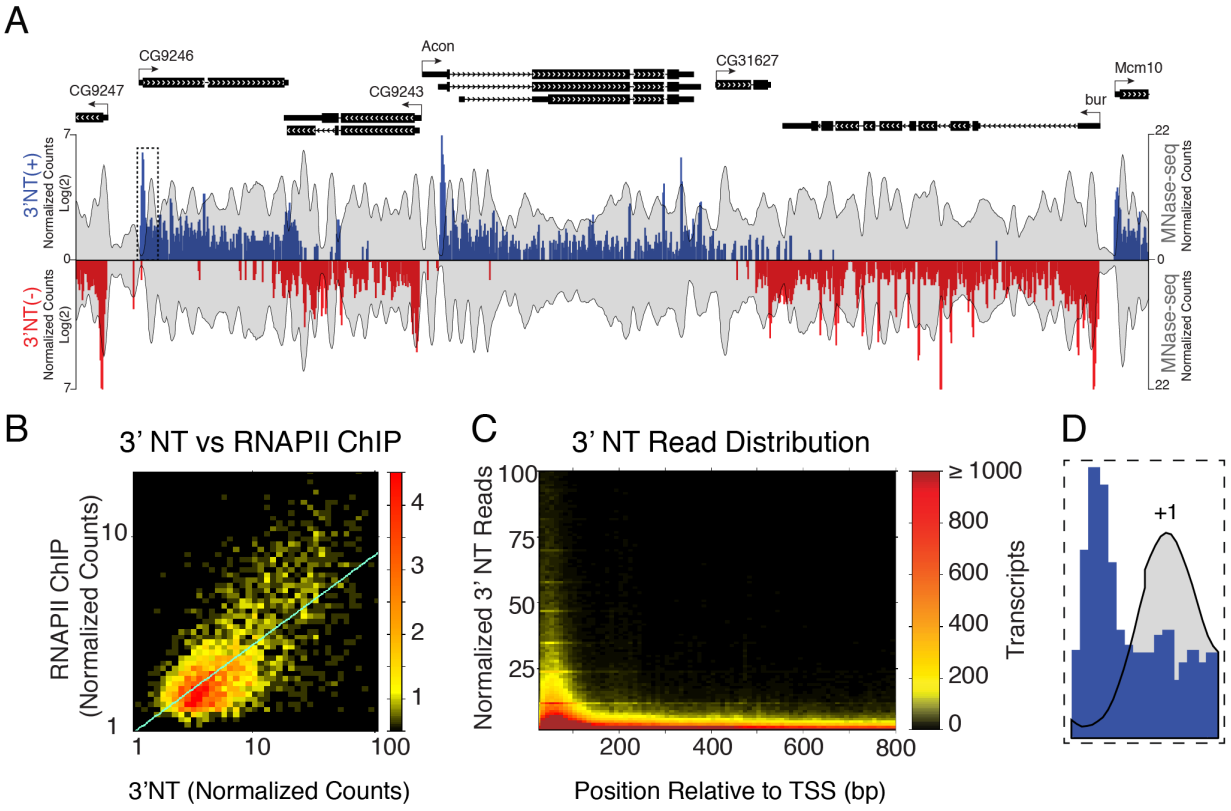


Figure 3-2. Precise strand-specific mapping of total RNAPII at nucleotide resolution. (A) A representative genome browser snapshot of highly expressed genes (Chr2L:21,162,558 - 21,180,500) showing the 3' position of nascent transcripts¬ (3'NT) that map to + (above axis, blue) and – (below axis, red) strands overlaid onto the nucleosome landscape map (mirrored, grey) derived from MNase-seq. (B) Log-log density plot comparing our 3'NT reads with reads from RNAPII ChIP-seq (r=0.69) (Core et al., 2012) aggregated from TSS to TSS+250 bp (normalized reads/bp). (C) 2D histogram of 3'NT for all transcripts in the Drosophila genome plotted as a heat map. All 3'NT and MNase-seq data are from two biological replicates. (D) Boxed region shown in (A) at higher magnification.

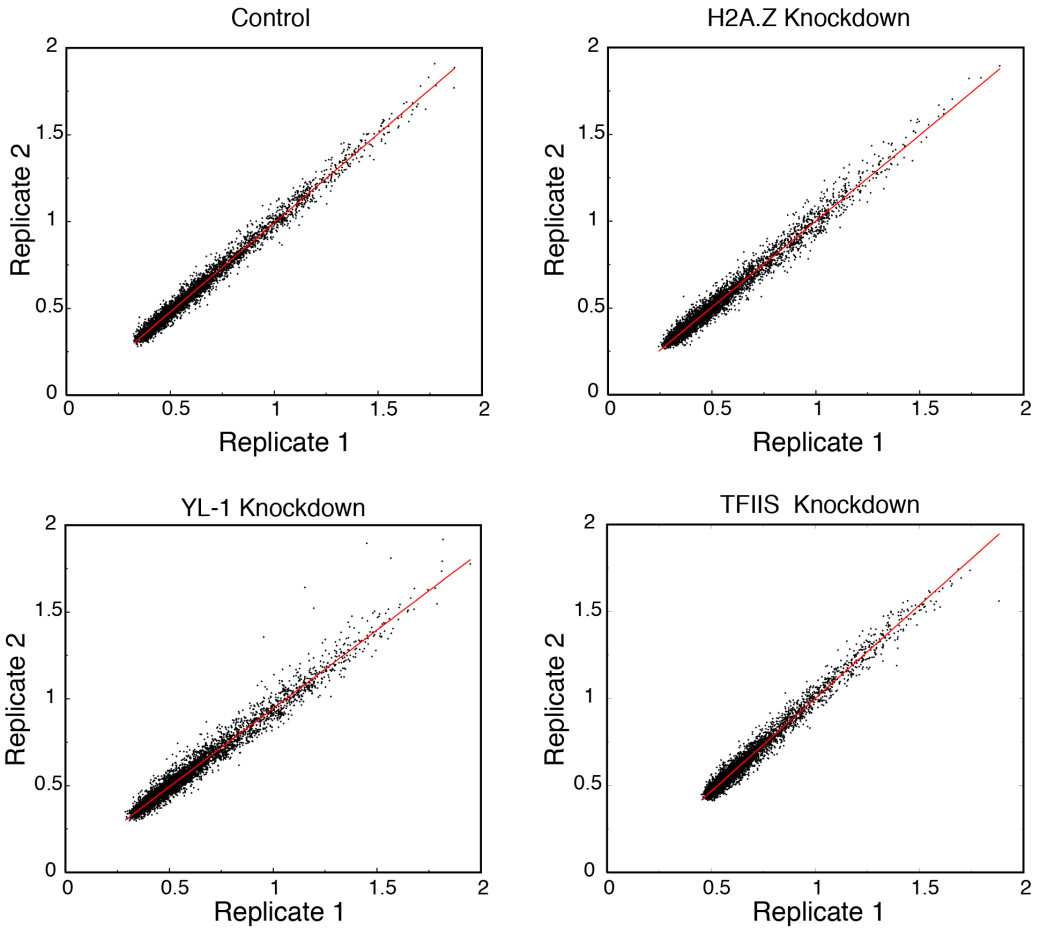


Figure 3-3. High correlation of 3'NT signal between biological replicates.
Scatter plots on log-log scale, comparing average reads per transcript of individual replicates from control (Pearson correlation, r=0.99), H2A.Z KD (r=0.99), YL1 KD (r=0.98), TFIIS KD (r=0.99). The linear fit comparing the replicates is shown as red line.

Summary Of 3'NT Reads

Sample	Reads Mapped to Genome	Total	Non-coding RNA reads filtered out
Control A	73,393,169	151,769,239	43,245,130
Control B	78,376,070		48,323,571
H2A.Z KD A	81,030,839	159,950,991	37,528,164
H2A.Z KD B	78,920,152		37,398,454
YL-1 KD A	77,462,080	151,737,588	39,735,215
YL-1 KD B	74,275,508		41,785,113
TFIIS KD A	49,134,049	103,877,292	29,999,938
TFIIS KD B	54,743,243		32,715,402

Table 3-1

Entry to the +1 nucleosome is a major barrier in vivo

To obtain a global view of 3'NT signal from all transcripts, we plotted a 2D histogram of the normalized reads with respect to the TSS (Figure 3-2C). The highest 3'NT signal for most transcripts is within ~150 bp of the TSS and trails off to a relatively uniform distribution over downstream gene bodies. In *Drosophila* this roughly corresponds to the vicinity of the +1 nucleosome, but the exact position is variable (Mavrich et al., 2008; Weber et al., 2010). When we overlaid the nucleosome landscape derived from paired-end sequencing of DNA fragments protected from micrococcal nuclease digestion in nuclei (MNase-seq), we observed that the 3'NT signal is highest just upstream of the +1 nucleosome (Figure 3-2A and 3-2D).

As the highest density of 3'NT signal was found around the +1 nucleosome boundary, we asked where RNAPII tracks with respect to the +1 nucleosome of each transcript. First, we defined the precise position of the +1 nucleosome dyad for each transcript from paired-end MNase-seq landscapes. We then grouped transcripts into quartiles based on 3'NT reads over the flanking region (2 kb downstream of the TSS) and ordered transcripts in each quartile by the distance between their TSS and the dyad axis of their +1 nucleosome. The called dyad axis position can be clearly visualized on the MNase-seq heat map (Figure 3-4A), validating the accuracy of the called position (Table 3-2, Figure 3-5A). We next plotted the 3'NT heat map with the same order of transcripts, which revealed a striking correspondence of 3'NT signal density with nucleosome positions. Specifically, the leading edge of the 3'NT signal tracked with the entry site of the nucleosome (dyad – 73 bp) for +1 nucleosome positions independent of their distance from the TSS. This trend was observed for all quartiles suggesting that this is an intrinsic effect of RNAPII encountering the +1 nucleosome, independent of the level of transcription and the nucleosome distance from the TSS.

In addition to the signal that aligns with the +1 nucleosome, we also observed 3'NT enrichment in the region between the TSS and +1 nucleosome entry site for many transcripts in the first quartile (Figure 3-4A, top). This high signal is likely due to promoter-proximal paused RNAPII (Adelman and Lis, 2012; Yamaguchi et al., 2013). To investigate this, we plotted the 3'NT signal for transcripts by degree of pausing, as defined by the change in RNAPII when negative elongation factor (NELF) is depleted (Gilchrist et al., 2010) (Figure 3-5B-C). We observed the highest 3'NT enrichment in the known region for paused polymerase at the most NELF-affected transcripts (~20-60 nucleotides from the TSS). However, the leading edge of the 3'NT signal also tracked with the +1 nucleosome for all NELF-affected quartiles, similar to our observation for all genes. This suggests that apart from being subject to NELF-regulated pausing just downstream of the TSS, RNAPII stalls when it encounters the downstream +1 nucleosome.

We next sought to identify the position at the +1 nucleosome that is most refractory to RNAPII transit. The single nucleotide resolution and high dynamic range of 3'NT enable us to observe individual spikes indicating consistent RNAPII stall sites. To characterize these positions, we define a stall as any single nucleotide spike that is at least 3 standard deviations higher than the mean in a 200-bp window. We first investigated the probability of stalling (stall density) at positions relative to the +1 nucleosome dyad for genes grouped according to 3'NT quartiles as before. Averaging over thousands of transcripts, we observed a defined stall density peak at -80 bp from the dyad (Figure 3-4B, see also Figure 3-5D). Considering the 15-20 bp distance between the active site and the leading edge of RNAPII (Samkurashvili and Luse, 1996), the -80 bp stall position corresponds to the RNAPII leading edge positioned at -65 to -60 bp from the dyad or ~8-13 bp within the nucleosome. We observed a similar peak for all quartiles of average 3'NT signal suggesting that this stall position is intrinsic to the physical

barrier of the nucleosome. Thus, based on 3'NT signal tracking with the +1 nucleosome entry site, and mapping of the stall sites, we conclude that entry to the +1 nucleosome is a major barrier to transcription.

RNAPII backtracks when it stalls at nucleosomes

In vitro, RNAPII has been shown to backtrack whenever there is an impediment to elongation (Komissarova and Kashlev, 1997; Nudler et al., 1997). If the entry site of the +1 nucleosome is a true barrier to RNAPII, then a fraction of RNAPII is expected to be backtracked, similar to observations *in vitro* (Kireeva et al., 2005). As our 3'NT signal maps both elongating and backtracked/arrested RNAPII, and PRO-seq captures only the elongating form, we can estimate the fraction of RNAPII that is backtracked by comparing 3'NT to published PRO-seq data for *D. melanogaster* S2 cells (Kwak et al., 2013). Because both methods entail quantitative recovery of nascent RNA and can be internally normalized for each transcript in the same way, we can subtract the PRO-seq enrichment at a given position from the 3'NT enrichment. If all bound RNAPII complexes are competent to elongate, we expect to see similar enrichment for 3'NT and PRO-seq. To estimate the extent of backtracking due to nucleosome barriers, we plotted 3'NT and PRO-seq relative to the dyad axis (Figure 3-4C, see also Table 3-3 and Figure 3-5E). We observed high enrichment up to -80 bp for 3'NT and PRO-seq, but 3'NT enrichment was ~1.6x higher. This suggests that a significant fraction of RNAPII is backtracked/arrested and unable to elongate in a run-on assay due to encountering the nucleosome barrier.

In order to directly determine whether stalled RNAPII that is unable to elongate in a run-on assay is backtracked, we used RNAi to deplete the transcript cleavage factor TFIIS from cells and mapped the resulting 3'NT positions. TFIIS stimulates RNA cleavage from backtracked

RNAPII complexes, assisting in their reactivation (Sigurdsson et al., 2010). TFIIS knockdown did not change the position of the major barrier location at the +1 nucleosome (-80 bp from the dyad), as expected because TFIIS acts only after RNAPII backtracks (Figure 3-5F-G). Furthermore, TFIIS knockdown is unlikely to affect nucleosome positioning, because TFIIS action is limited to its effect on RNAPII, which has been shown in yeast and mammalian cells not to alter nucleosome positioning (Kouzine et al., 2013; Zhang et al., 2011). We have confirmed these results for Drosophila S2 cells, by showing that inhibition of transcriptional elongation has no effect on nucleosome positioning (Figure 3-5H).

TFIIS knockdown changed where RNAPII stalls, such that stalling decreased at the entry site and increased within the +1 nucleosome, closer to the dyad (Figure 3-4D and 3-5G). This increase within the nucleosome indicates that RNAPII unwraps the nucleosome to various degrees before backtracking to the nucleosome entry site, before RNA cleavage. Thus, RNAPII backtracks due to encountering a nucleosome barrier.

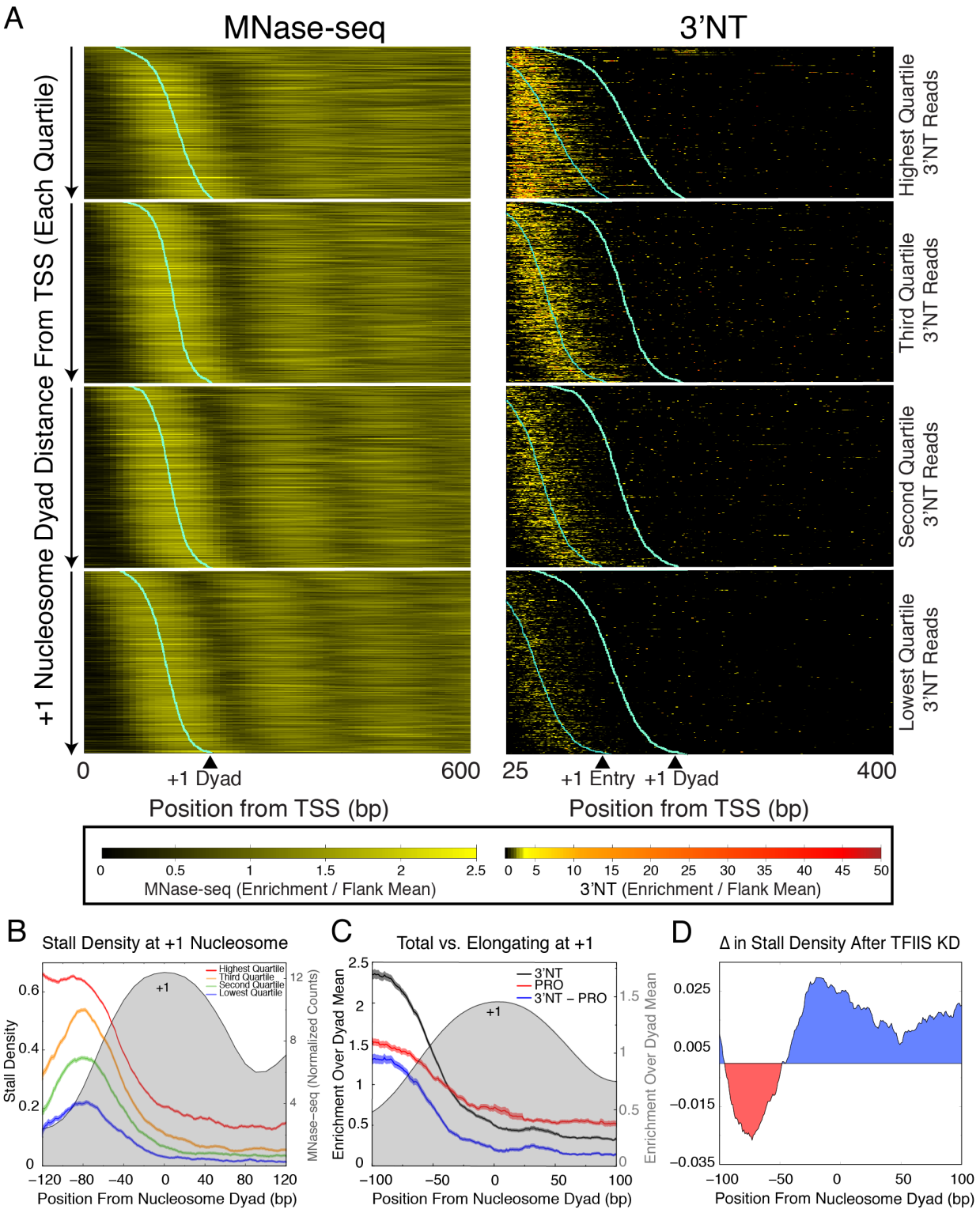


Figure 3-4. Entry to the +1 nucleosome is a major barrier *in vivo.*
(A) Transcripts (n=3981) were grouped into expression quartiles based on average 3'NT reads in the downstream flanking region (2 kb downstream of the TSS) and ordered from highest (top) to lowest quartile (bottom). Within each quartile, transcripts were arranged by distance between

the TSS and the +1 nucleosome dyad axis (shortest to longest). Data within each quartile are shown as enrichment over the downstream flank mean (2 kb downstream of the TSS) for MNase-seq (left) and 3'NT (right). The called +1 dyad axis position for each transcript is overlaid onto normalized data for both MNase-seq & 3'NT. The calculated entry site (dyad axis position – 73 bp) is also shown for 3'NT (right). (B) Stall Density at the +1 nucleosome, grouped into expression quartiles based on the downstream flank mean (2 kb downstream of the TSS) of the transcripts, overlaid onto the +1 nucleosome (grey) from MNase-seq. (C) Comparison between total RNAPII (3'NT) and elongation competent RNAPII (PRO = PRO-seq) for transcripts (n=715) that met the coverage criterion in PRO-seq with respect to the dyad axis of the +1 nucleosome (grey) with plots of mean ± Standard Error of the Mean (S. E. M.) (D) Change in stall density at the +1 nucleosome after depletion of TFIIS by RNAi.

Nucleosomes Called Within Each Expression Quartile

Quartile	# of Transcripts	Called +1	Percent Called
1 (lowest)	1170	1038	88.7
2	1170	1042	89.1
3	1170	1032	88.2
4 (highest)	1170	869	74.3
Total	4680	3981	85.1

Table S3. Comparing PRO-seq to 3'NT

# of Transcripts with coverage cutoff used for 3'NT	Called +1	Percent Called
715	564	78.9

Table 3-2

Summary of Nucleosomes Called Per Position

Nucleosome Position	Nucleosomes Called
1	3514
2	1636
3	1161
4	933
5	771
6	578
7	456
8	321
9	219

Table 3-3

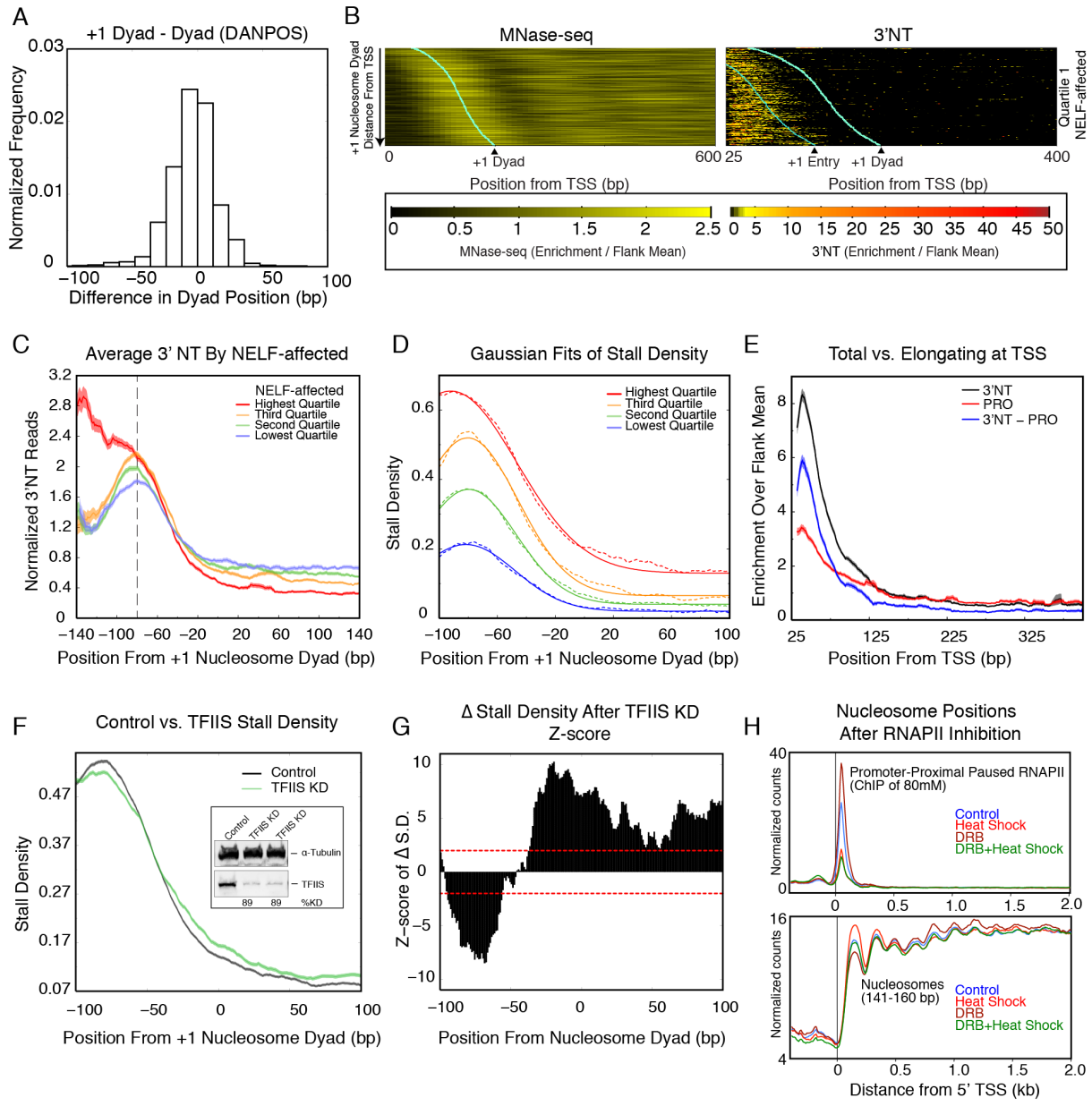


Figure 3-5

(A) Calling of dyad position correlates highly with the position called by DANPOS, shown as the difference in dyad position (Chen et al., 2013a). (B) 3'NT signal tracks with the entry site to the +1 nucleosome at most NELF-affected transcripts similarly to all genes. Most NELF-affected genes (Quartile 1) with a called +1 nucleosome dyad (n=908) were arranged by distance between the TSS and the +1 nucleosome dyad axis (shortest to longest). Data are shown as enrichment over the downstream flank mean (2 kb downstream of the TSS) for MNase-seq (left) and 3'NT (right). The called +1 dyad axis position for each transcript is overlaid onto normalized data for both MNase-seq & 3'NT. The calculated entry site (dyad axis position - 73 bp) is also shown for 3'NT (right). (C) The +1 barrier and leading edge is highly similar for all NELF-affected quartiles. Average 3' NT normalized reads by NELF-affected quartile, with plots of mean ± Standard Error of the Mean (S. E. M.) with respect to the +1 nucleosome dyad. Most NELF-

affected quartile shows high 3'NT signal upstream of the nucleosome as expected for the pause site, whereas the downstream leading edge tracks at the -80bp nucleosome barrier location for all quartiles. The -80 bp position is marked by a dashed line. (D) Gaussian fit (solid line) of the Stall Density (dashed line) at the +1 nucleosome for different expression quartiles with mean positions at -81.5, -79.7, -80.4, and -91.9 for lowest, second, third and highest quartiles respectively. Highest quartile displays a slight upstream shift because this quartile contains many paused genes with high 3'NT signal close to the promoter. (E) A significant fraction of RNAPII is not competent to elongate downstream of the TSS. Comparison between total RNAPII (3'NT) and elongation competent RNAPII (PRO = PRO-seq) for 715 transcripts that met the coverage criterion in PRO-seq relative to the TSS with plots of mean ± Standard Error of the Mean (S. E. M.). (F) Average stall density with respect to the +1 nucleosome dyad for control and after TFIIS knockdown, showing decrease upon entry and increased internal stalling. Error bars calculated according to binomial distribution. (inset) Western blot from whole cell extracts show efficient TFIIS knockdown for two biological replicates. (G) Z-score calculation of the difference in stall density at the +1 nucleosome after depletion of TFIIS by RNAi. Red line indicates z-score = 2 & -2 ($p < 0.05$). (H) (Top) Promoter-proximal paused RNAPII (ChIP from small fraction solubilized in 80mM NaCl after MNase digestion) becomes highly enriched upon DRB treatment and is strongly reduced upon heat shock (Bottom) Nucleosome positions do not change after RNAPII inhibition with DRB treatment, heat shock, or both treatments together (Teves and Henikoff, 2011).

The nucleosome barrier is context-specific

The +1 nucleosome position is different from gene body nucleosomes because it is flanked by a nucleosome on only the downstream side. On the upstream side there is a nucleosome-depleted region, where the transcription machinery assembles before extending into productive elongation. The average distance between the TSS and the entry site of the +1 nucleosome (57 bp) is also considerably longer than the average length of linker DNA between downstream nucleosomes (20 bp). Hence, we asked if this distinction might also confer differences in how RNAPII transcribes through downstream nucleosomes relative to the +1 nucleosome. We found that the major barrier position at the +2 nucleosome was highly similar to that of the +1 nucleosome position (~ -80 bp), but with lower magnitude, and trailed slightly further into the nucleosome (Figure 3-6A). Thus, the barrier at -80 bp that we observed for the +1 nucleosome is not strictly due to being the first barrier to transcription or an effect of the

promoter, as we observed this barrier position at the +2 nucleosome also. However, nucleosomes farther downstream displayed a barrier that is distinct from the +1 and +2 nucleosomes. In these gene body nucleosomes (> +2), RNAPII stalls upon entry (-80 bp), similar to the +1 and +2 nucleosomes, but also before the dyad axis (~ -50 & -20 bp). A similar trend was observed when we plotted the average of the 3'NT reads for these positions (Figure 3-7). Figure 3-6B summarizes the relative magnitude and the locations of the stalls relative to the leading edge of RNAPII's (active site +20 bp) position on the nucleosome. Thus, nucleosomes present unique stall signatures depending on their location downstream of the TSS.

Considering differences in how RNAPII transcribes through +1, +2 and gene body nucleosomes, we asked if the magnitude of the barrier is different at different nucleosome positions (Table S4). To measure the barrier, we calculated the fraction of positions that have a stall from 100 bp upstream of the dyad to the dyad (stall fraction, see Methods). We also calculated the average 3'NT signal from 100 bp upstream of the dyad up to the dyad. We found that the +1 nucleosome position had the highest stall fraction and average 3'NT signal ahead of the dyad indicating a higher barrier, whereas the downstream nucleosome positions were much less of a barrier and had relatively uniform magnitude (Figure 3-6C). Additionally, the higher stall fraction at +1 was highly similar for all +1 distances from the TSS, confirming that the increased barrier is due to the nucleosome and not skewed by proximity to the promoter (Figure 3-6D). These results suggest that once RNAPII clears the first major barrier, downstream elongation is less impeded, despite differences in the predominant location of the barrier within the nucleosome. We conclude that nucleosomes present a context-specific barrier to RNAPII, one in which nucleosome location relative to the TSS predicts the magnitude and characteristics of the barrier.

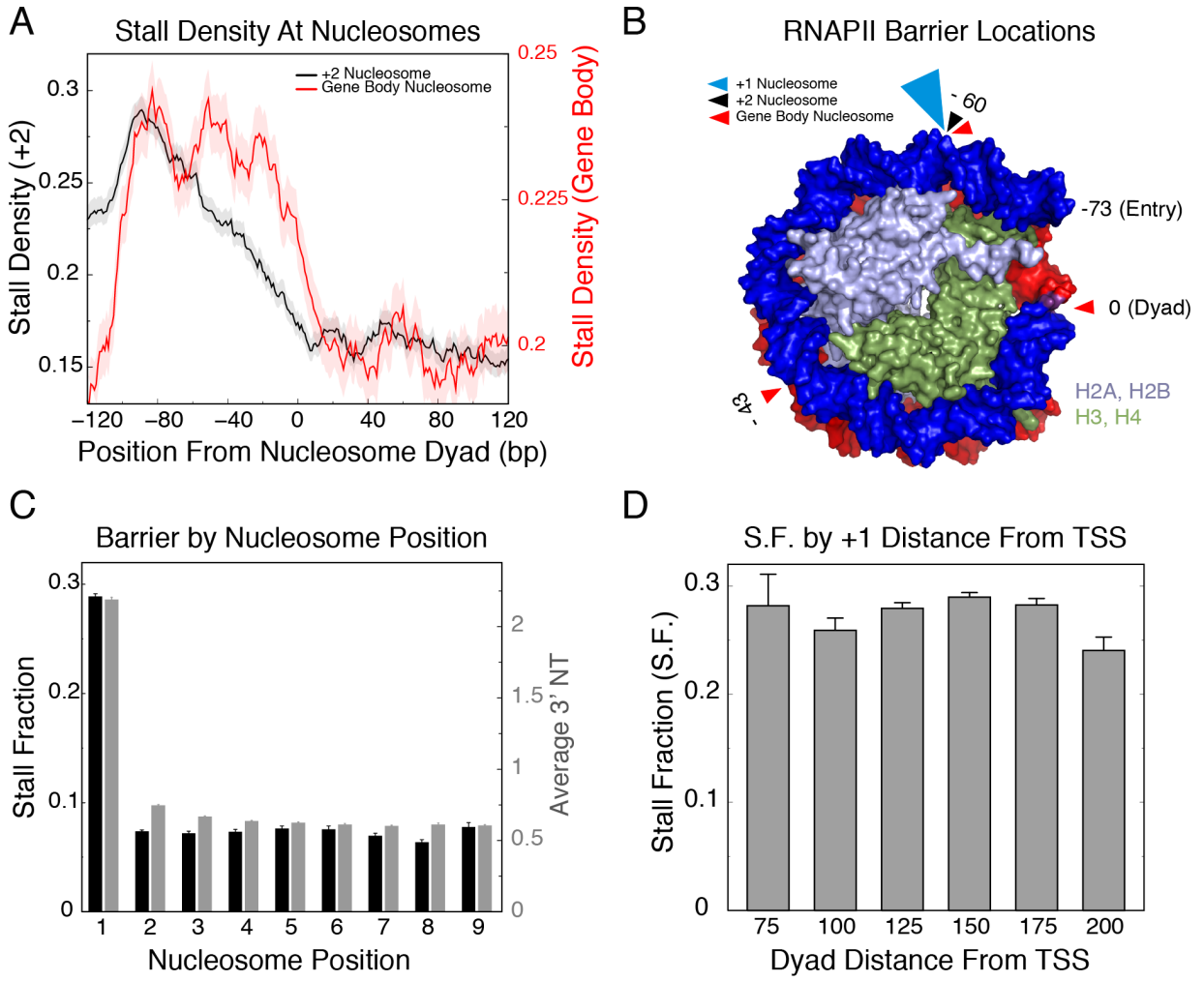


Figure 3-6. The nucleosome barrier is context-specific.
(A) Comparison of RNAPII Stall Densities at +2 and gene body nucleosomes. (B) Nucleosome structure with marked barrier locations for the leading edge of RNAPII that first encounters the nucleosome (barrier position detected by 3'NT + 20 bp). Size of arrowhead(s) reflects cumulative barrier magnitude relative to +1. (C) Stall Fraction and average 3'NT signal ahead of the dyad axis at nucleosome positions 1-9. (D) Stall Fraction (S.F.) is determined by the nucleosome and not proximity to the promoter as S.F. is uniformly high for all dyad distances from the TSS.

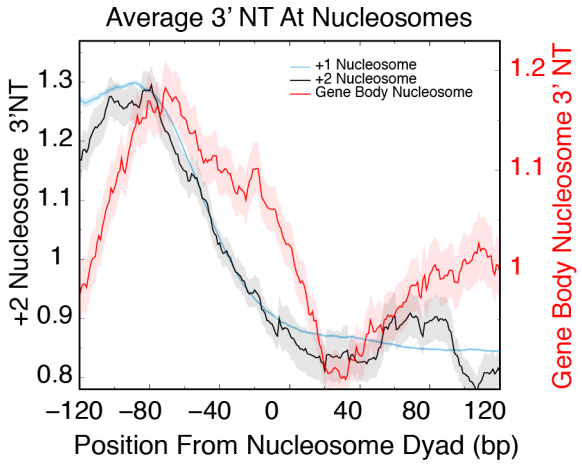


Figure 3-7
Average 3' NT signal at +2 and Gene Body (G.B.) with plots of mean ± Standard Error of the Mean (S. E. M.), resembles trend observed by stall density. Note that the +1 nucleosome data is scaled down for comparison to +2 and G.B.

Nucleosomes determine the degree of RNAPII stalling

If the +1 nucleosome indeed presents a physical barrier to RNAPII that leads to a defined stall position, then the degree of stalling should reflect nucleosome occupancy. To test this, we grouped transcripts based on the stall fraction and plotted the nucleosome occupancy, as determined by MNase-seq, with respect to the different stall fraction magnitudes. We observed a trend of increased +1 nucleosome occupancy with increased stall fraction (Figure 3-8A), indicating that nucleosome occupancy defines the extent of RNAPII stalling.

A unique feature of the +1 nucleosome and a few nucleosomes downstream is the conspicuous enrichment of H2A.Z (Mavrich et al., 2008; Weber et al., 2010). H2A.Z is incorporated into nucleosomes around promoters of active genes independent of replication, suggesting that H2A.Z might influence transcription (Cairns, 2009). Hence, we investigated H2A.Z occupancy of genes grouped according to stalled fraction. If H2A.Z does not influence RNAPII transit, then we expect genes with high +1 nucleosome occupancy to also have high H2A.Z and correlate with the extent of RNAPII stalling. In stark contrast, we find that absolute

H2A.Z levels *anti*-correlate with nucleosome occupancy and RNAPII stalling, such that genes with a lower stall fraction have higher H2A.Z levels and *vice versa* (Figure 3-8A). This suggests that where the extent of RNAPII stalling is less, H2A.Z is disproportionately increased.

The opposing trends of nucleosome occupancy and H2A.Z enrichment on stalling can also be seen by plotting average MNase-seq and H2A.Z ChIP-seq signal relative to the TSS for genes grouped according to stall fraction (Figure 3-8B). This trend is also clearly exemplified in individual gene tracks (Figure 3-8C), where high nucleosome occupancy and low H2A.Z lead to more stalled positions. Thus, increased nucleosome occupancy corresponds to increased RNAPII stalling at the +1 nucleosome, while increased H2A.Z occupancy correlates with decreased stalling, suggesting that the presence of H2A.Z eases the inhibitory effect of nucleosomes on transcription.

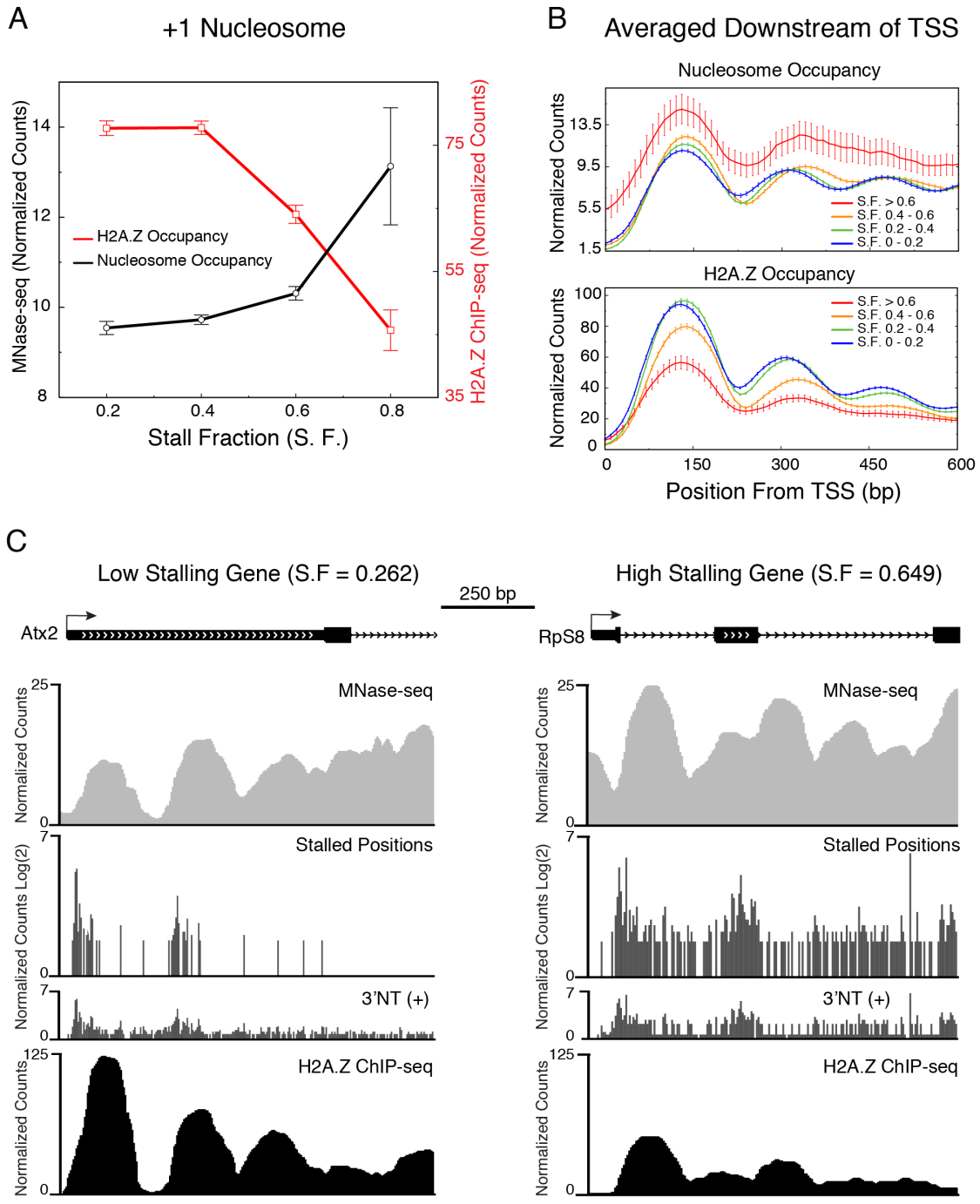


Figure 3-8. Nucleosomes determine the extent of RNAPII stalling.
(A) +1 Nucleosome and H2A.Z occupancy at genes with different Stall Fraction (S.F.) magnitudes derived from the +1 nucleosome position (n=3981). (B) MNase-seq and H2A.Z ChIP-seq normalized counts averaged over genes grouped by Stall Fraction (derived from +1 position) plotted relative to the TSS. Error bars represent S. E. M. (C) UCSC Genome Browser snapshots for two genes with low stalling (left), showing low nucleosome occupancy (MNase-seq, top track) and high H2A.Z (bottom track), and high stalling (right), showing high nucleosome occupancy (above) and low H2A.Z (below), with the 3'NT normalized counts and stalled positions shown in between.

H2A.Z modulates the nucleosome barrier to RNAPII

Our observations point to a role for H2A.Z in reducing the nucleosome barrier to RNAPII. To test this hypothesis, we depleted H2A.Z from chromatin using RNA interference (RNAi) against H2A.Z and YL-1. YL-1 is orthologous to budding yeast Swc2, a subunit of the Swr1 complex that binds H2A.Z and is essential for ATP-dependent exchange of H2A with H2A.Z (Wu et al., 2005). H2A.Z depletion by impairing Swr1-dependent incorporation provides independent confirmation of direct H2A.Z knockdown and controls for potential phenotypes caused by Swr1 activity alone (Halley et al., 2010). Within the chromatin fraction, H2A.Z knockdown reduced protein levels to 12% and 13% of control and YL-1 knockdown reduced H2A.Z levels to 64% and 62% for the two biological replicates (Figure 3-9A).

We then determined the genome-wide occupancy of the remaining H2A.Z by ChIP-seq and mapped the 3'NT positions in the knockdown cells. Strikingly, nucleosome positions with increased RNAPII stalling after H2A.Z knockdown had significantly less H2A.Z relative to positions where stalling was unchanged, for +1, +2, and gene body nucleosomes (Figure 3-9B, see also Figures 3-10 and 3-11). Importantly, this effect was graded in the YL-1 knockdown due to less efficient H2A.Z depletion, leading to smaller effects at the +1 and +2 nucleosome positions. GO analysis shows that metabolic processes and catalytic activity genes were over-representated ($p < \sim10^{-2}$ to 10^{-6}), as expected for the S2 cell line. These results indicate that an important function of H2A.Z is to modulate the elongation kinetics of RNAPII both at the +1 nucleosome and further into gene bodies, which helps to explain its enrichment around promoters in virtually all eukaryotes.

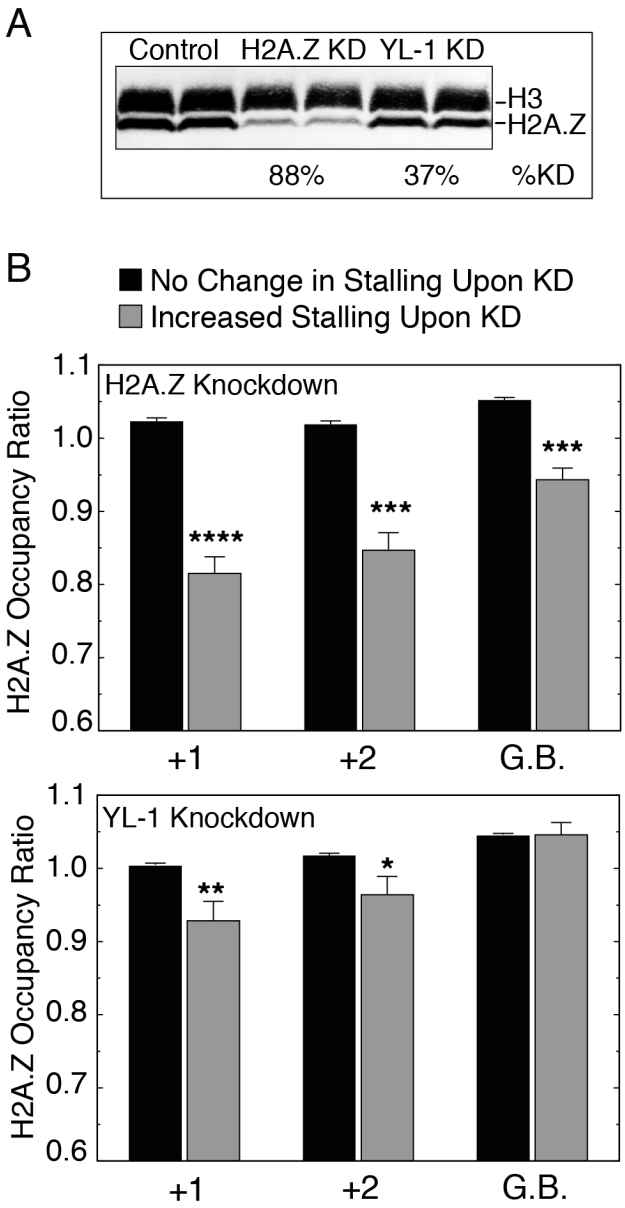


Figure 3-9. H2A.Z modulates the nucleosome barrier to RNAPII.
A) Western blot showing the extent of H2A.Z depletion from chromatin after H2A.Z and YL-1 KD. B) Nucleosome positions with increased RNAPII stalling defined by 3'NT after knockdown have significantly less H2A.Z relative to unchanged positions (H2AZ KD: +1 n=279, +2 n=145, G.B n=415, YL1 KD: +1 n=124, +2 n=98, G.B. n=260). Asterisks represent statistically significant differences.

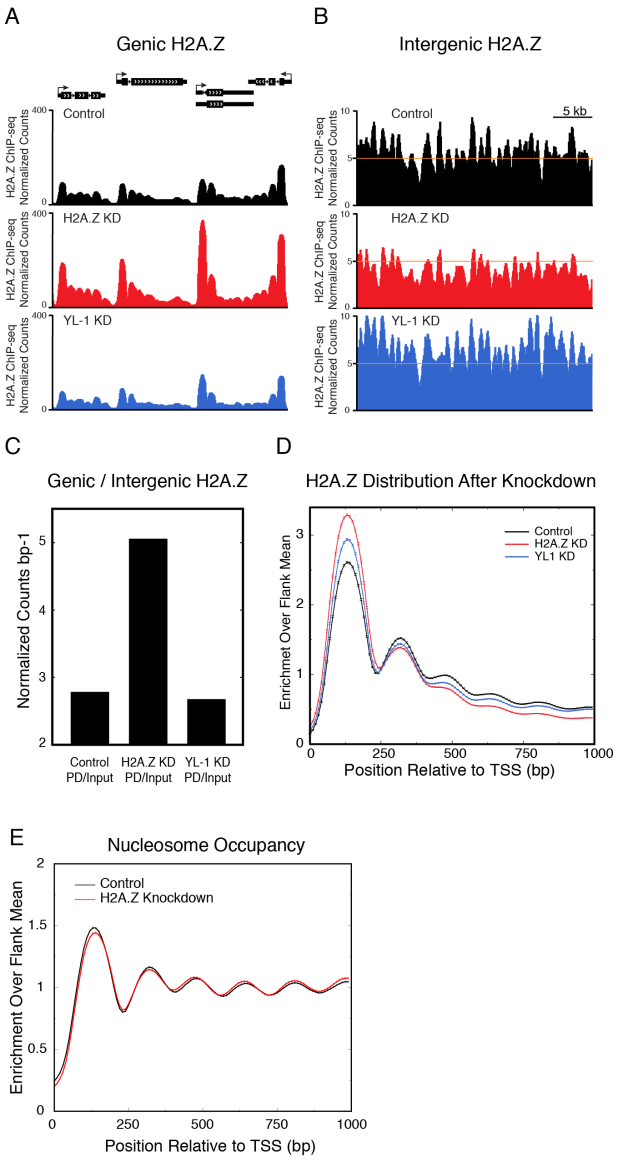


Figure 3-10. H2A.Z is globally redistributed after knockdown.
After H2A.Z knockdown, H2A.Z enrichment relative to the rest of the genome was higher at the +1 nucleosome and over gene bodies but decreased in intergenic regions, relative to control. YL-1 knockdown led to very subtle decrease in H2A.Z levels over gene bodies but was increased in intergenic regions. Representative UCSC Genome Browser snapshots showing relative differences in H2A.Z occupancy after knockdown (KD) over genes (A) (chr2L:13,366,500 - 13,371,000) and in an intergenic region (B) (chr2R:4,182,500 - 4,212,500), orange line is drawn at an arbitrary position for ease of comparison. (C) Genome-wide H2A.Z redistribution after KD, ChIP (pull-down (PD)/input) normalized counts from (TSS to TSS + 2kb)/(intergenic) per bp in control, H2A.Z or YL-1 KD (D) H2A.Z distribution after knockdown from the TSS. Data are plotted as enrichment over flanking region (1 kb downstream of the TSS), with plots of mean ± S. E. M. (E) Nucleosome occupancy is not significantly affected due to H2A.Z knockdown. Average MNase-seq profile plotted from the TSS as enrichment over flanking region (1 kb downstream of the TSS), with plots of mean ± S. E. M.

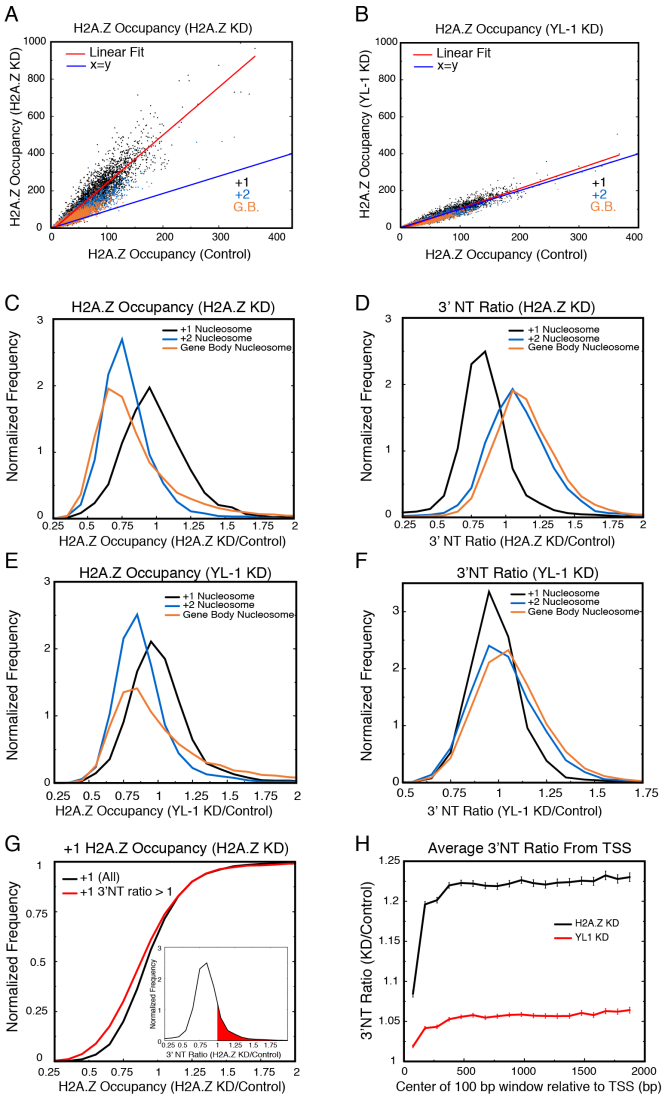


Figure 3-11. H2A.Z modulates the nucleosome barrier to RNAPII.
For both H2A.Z and YL-1 knockdown cells, loss of H2A.Z at a nucleosome position results in a higher barrier to RNAPII and conversely, gain of H2A.Z results in a lower barrier. Relative to +1 position, +2 and gene body nucleosomes have greater loss of H2A.Z upon knockdown of both H2A.Z and YL1 and conversely, the effect on RNAPII stalling is higher at +2 and gene body nucleosomes compared to +1. (A-B) Change in H2A.Z occupancy at +1, +2, and gene body (G.B.) nucleosomes after knockdown (KD) for H2A.Z KD (A) and YL-1 KD (B) cells. The linear fit for the +1 nucleosome and the line representing x=y are shown. (C) Distribution of the H2A.Z occupancy ratio (H2A.Z KD / Control) for +1, +2, and gene body nucleosomes after H2A.Z knockdown (KD). (D) Distribution of the 3'NT ratio (H2A.Z KD/Control) for +1, +2, and gene body nucleosomes. (E-F) Same as C-D, respectively, but for YL-1 KD cells. (G) Cumulative distribution of H2A.Z occupancy (H2A.Z KD / Control) for the +1 nucleosome plotted for all transcripts (black) and the subset of transcripts with a 3'NT ratio greater than 1 (red). (inset) Distribution of the 3'NT ratio of +1 nucleosome, where the region with 3'NT ratios > 1 is shaded red. (H) Average 3'NT ratio in 100-bp windows with centers at TSS+75, TSS+175, TSS+275 and so on, plotted with mean ± S. E. M. for H2A.Z KD (black) and YL-1 KD (red).

H2A.Z occupancy anti-correlates with H3-H4 nucleosome turnover

How might H2A.Z modulate RNAPII kinetics? We hypothesized that H2A.Z might make the nucleosome easier to disrupt when encountered by RNAPII. We tested this by measuring the nucleosome turnover defined by CATCH-IT (Deal et al., 2010; Teves and Henikoff, 2011) to see if turnover correlated with H2A.Z occupancy. Much to our surprise, we found the opposite, that the level of H2A.Z occupancy anti-correlated with complete nucleosome turnover (Figure 3-12A). CATCH-IT measures nucleosome turnover as the replacement of H3 and H4, but not H2A or H2B, which suggests that H2A.Z-H2B dimers facilitate retention of $(H3\text{-}H4)_2$ tetramers.

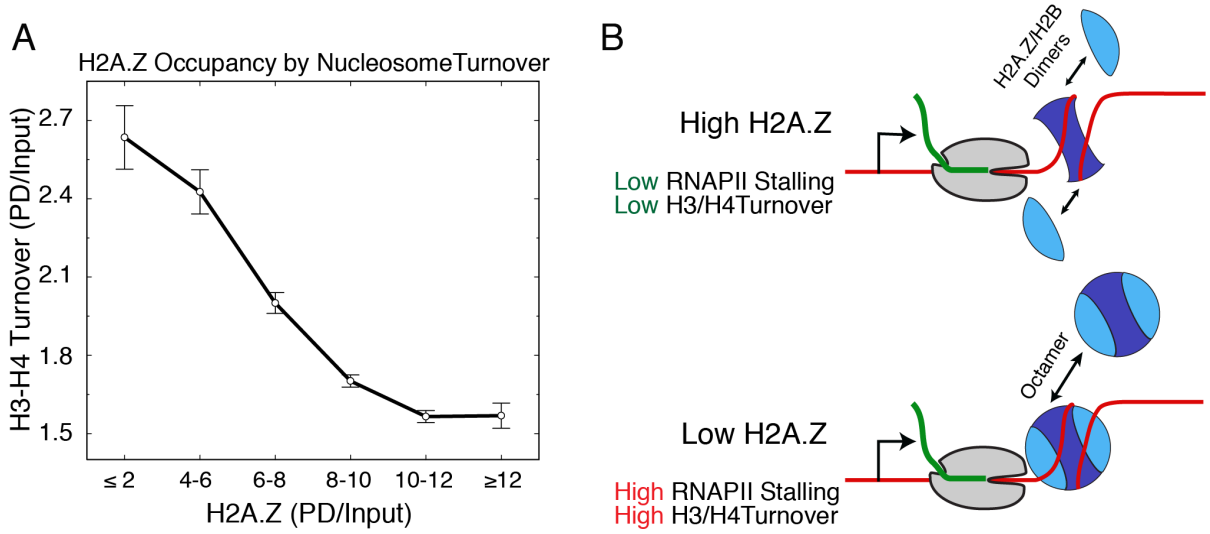


Figure 3-12. H2A.Z occupancy anti-correlates with H3-H4 nucleosome turnover.
A) H2A.Z occupancy anti-correlates with turnover of H3-H4 defined using CATCH-IT. (B) Model describes a likely mechanism for how H2A.Z reduces the barrier to RNAPII.

Discussion

Chromatin is the in vivo template for eukaryotic transcription, however the fundamental process whereby RNAPII transcribes across a nucleosome is poorly understood. In vitro, RNAPII is known to behave as a Brownian ratchet, stalling at specific locations associated with strong histone-DNA contacts (Bintu et al., 2012; Selth et al., 2010) whereas *in vivo*, a lack of suitable techniques has hampered understanding of this process. Using a simple technique that comprehensively maps total RNAPII at base-pair resolution in metazoan cells, we have gained fundamental insights into regulation of RNAPII elongation by nucleosomes. Our genome-wide map of the base-pair position at which the last RNA nucleotide is incorporated represents the ultimate metric of the process of transcription.

In vitro, the precise barrier position on a nucleosome template can be determined by the length of RNA produced (Kireeva et al., 2005). Under permissive conditions, a fraction of RNAPII can transcribe through the nucleosome, but smaller length RNAs are produced when RNAPII encounters an insurmountable mechanical barrier, backtracks and arrests. There is a consensus from a large body of work that +15 and +45 bp into the nucleosome are the major barrier positions (Bondarenko et al., 2006; Kulaeva et al., 2013). The leading edge of RNAPII encounters the nucleosome 15-20 bp in front of the active site where polymerization takes place (Samkurashvili and Luse, 1996), whereby the leading edge of RNAPII is around +35 and +65 bp into the nucleosome. These positions have been attributed to strong histone-DNA interactions and have been observed even in the presence of high salt and elongation factors such as TFIIS and FACT. Surprisingly, our comprehensive single nucleotide resolution maps of RNAPII *in vivo* reveal nucleosomal barriers that differ from the *in vitro* consensus. The consensus barrier position we find for all nucleosome positions is at -7 bp from the nucleosome entry site (-80 bp

from the dyad axis), putting the leading edge of RNAPII at just +8 to +13 bp into the nucleosome, approximately a single helical turn of DNA. It is intriguing that the nature of the nucleosome barrier *in vivo* is so distinct from that *in vitro*. This distinction might be attributed to torsional constraint of the template *in vivo* but not *in vitro*, where the accumulation of positive supercoils ahead of RNAPII destabilizes nucleosomes (Sheinin et al., 2013; Teves and Henikoff, 2014). Hence, our data point to a different mechanism of RNAPII transit across nucleosomes *in vivo* compared to existing models using purified components and reconstituted nucleosomes.

Our analysis *in vivo* allows investigation into different nucleosome positions with respect to the TSS. We found that the overall magnitude of stalling is highest at the +1 nucleosome. Gene body nucleosomes also contain stalling positions further into the nucleosome, more closely resembling what has been observed *in vitro* (Kulaeva et al., 2013) and in budding yeast (Churchman and Weissman, 2011), albeit at much lower magnitude than is observed for the +1 nucleosome. It is possible that the relative lack of H2A.Z and/or histone modifications enriched on the +1 and +2 nucleosomes (Zentner and Henikoff, 2013) account for the closer resemblance of RNAPII transit at gene body nucleosomes to its action on unmodified canonical nucleosomes *in vitro*.

Arrested and/or backtracked RNAPII complexes detected by 3'NT accounts for ~60% of the total 3'NT signal at the nucleosome (Figure 3-4C) and are distinct from RNAPII that is found at the promoter-proximal paused position immediately downstream of the TSS. We observed the expected signal for paused RNAPII close to TSS, but we also observed RNAPII tracking with the nucleosome, similar to non-paused genes. Thus, while an effect of the promoter sequence on pausing is clear (Kwak et al., 2013; Lagha et al., 2013), our results suggest that after RNAPII pause release, the complex encounters a nucleosome barrier, which might contribute to

promoter-proximal pausing (Mavrich et al., 2008). In addition 3'NT detected barriers with a similar stall signature at downstream nucleosomes, where promoter-proximal pausing is not a complicating factor.

The +1 and +2 positions are much more highly enriched than gene body nucleosomes for H2A.Z, which is extensively acetylated on the N-terminal tail (Bruce et al., 2005; Valdes-Mora et al., 2012) and stimulates chromatin remodeler activity (Goldman et al., 2010). Wherever the barrier to RNAPII was increased after RNAi, we observed that H2A.Z levels were significantly decreased relative to unchanged positions. These observations are supported by results in yeast, where H2A.Z increased the elongation rate of RNAPII at a single fusion gene (Santisteban et al., 2011), and indicate that higher H2A.Z levels in gene bodies might influence expression of responsive genes (Coleman-Derr and Zilberman, 2012).

We were surprised by evidence suggesting that H2A.Z preserved $(H3\text{-}H4)_2$ tetramers during transcription but this provided mechanistic insight how H2A.Z might modulate the nucleosome barrier. Whereas many studies have investigated the stability of the H2A.Z containing nucleosome, it remains unresolved whether H2A.Z stabilizes or destabilizes the intact nucleosome (Bonisch and Hake, 2012). However, *in vivo* results suggest that at dynamic regions of the genome where nucleosomes are disrupted, H2A.Z containing nucleosomes are especially labile (Jin and Felsenfeld, 2007). The crystal structure of the H2A.Z nucleosome showed that a single amino acid difference in H2A.Z compared to H2A results in destabilization of its interaction with H3-H4 through its docking domain (Suto et al., 2000). We propose that when RNAPII enters the nucleosome, slightly unwrapping it, the H2A.Z-H2B dimer will be more easily lost, enhancing the elongation of RNAPII through nucleosomes (Figure 3-12B). By our model, when H2A.Z is low, RNAPII will stall more because the entire nucleosome must

sometimes be removed. However, we cannot rule out the alternative possibility that $(H3\text{-}H4)_2$ tetramers with lower turnover rates are more likely to retain H2A.Z. In either case, the effect of H2A.Z manifests both at the +1 nucleosome and further into gene bodies and suggests that H2A.Z is most enriched where the barrier is largest, to assist in transcriptional elongation.

Methods

Cell Culture and RNAi

Drosophila S2-DRSC cells were obtained from the Drosophila Genomics Resource Center (Stock #181). Cells were cultured in Schneider's media supplemented with 10% heat inactivated FBS at 25°C. dsRNA was synthesized using NEB reagents and protocols from PCR templates containing the T7 promoter sequence. PCR primers were as follows: H2Av Forward (5'-TAATACGACTCACTATAGGGCGAAACCGAATTCCGTAGAA - 3') H2Av Reverse (5-TAATACGACTCACTATAGGGAGTAGGCCTGCGACAGA -3'), YL-1 (DRSC02765), GFP control (Hamada et al., 2005), TFIIS (Adelman et al., 2005). To administer dsRNA, 2.5 x 106 log phase cells / cm2 surface area were brought up in Schneider's without FBS (0.1mL/cm2) and seeded to 25, 75, or 150cm2 flasks. 30µg dsRNA/1x106 cells was added to cells for 1 hour with intermittent mixing, an equal volume of Schneider's with 20% FBS was added, and cells were grown for 96 hours before harvesting.

Nascent RNA isolation

We modified protocols from (Nechaev et al., 2010; Nishino et al., 2012; Wysocka et al., 2001) to fractionate cells and isolate nascent RNA. dsRNA treated cells from 150cm2 flasks were harvested and pelleted by centrifugation for 1' at 1,000 x g at 4°C. Cells were quickly washed

with 2mL ice cold PBS supplemented with 10μg/mL alpha-Amanitin to inhibit transcription, split into two separate 2ml tubes, and then pelleted by centrifugation 3' at 3,000 x g. The cells were then lysed in 0.9mL ice cold Buffer A (1μl/mL SUPERase•In, protease inhibitor (Roche), 10mM Hepes 7.9, 10mM KCl, 3mM MgCl2, 0.34M sucrose, 10% glycerol, 1mM DTT, 0.1% Triton X-100) supplemented with 10μg/mL alpha-Amanitin, 8' on wet ice with intermittent mixing. Nuclei were then pelleted as before and brought up in 0.9mL buffer B (1μl/mL SUPERase•In, 9mM EDTA, 0.2mM EGTA, 1mM DTT, protease inhibitor (Roche), 0.1% Triton X-100) 15' on wet ice, and repeated 2x or until nuclei lyse. The chromatin pellet was then brought up in 0.9mL buffer B+ (1μl/mL SUPERase•In, 20mM EDTA, 0.2mM EGTA, 2mM spermine, 5mM spermidine, 1mM DTT, protease inhibitor (Roche)) and washed 3x. The final insoluble pellet was brought up in 1mL Buffer B+ with 1% SDS and aliquots were harvested to assay knockdown efficiency or for western analysis. Antibodies for western blots are as follows: RNAPII (ab817), phospho S2 RNAPII (ab5095), H3 (ab1791), H2Av (Active Motif #39715), TFIIS (Adelman et al., 2005). The two pellets from the same original flask were pooled and dissolved in 10mL Trizol and RNA was isolated according to manufacturer's instructions. All RNA precipitation steps were conducted after incubation with 0.3M NaOAc and GlycoBlue. Biological replicates were collected for each treatment.

Solexa Sequencing and Alignment

3'NT library preparation

Nascent RNA was treated with DNase (RQ1, Roche) and extracted with acid phenol and ethanol precipitated with GlycoBlue and 0.3M NaOAc. We selected for the presence of a 5' cap using Terminator 5'-Phosphate-Dependent Exonuclease (Epicentre), as described (Nechaev et al.,

2010). Nascent, capped RNA was purified with acid phenol as before. Libraries were constructed as described for NET-seq (Churchman and Weissman, 2012) with the following modifications. Ligated RNA was digested with 2x alkaline fragmentation buffer at pH 10.3 for 30'. Reverse transcriptase primer concentration was reduced to 1.09 μM in RT reaction mix. Libraries were amplified with KAPA HiFi under limiting cycles. Following amplification, PCR reactions were purified with AMPure XP beads at a 1:1 ratio before PAGE purification to aid in the elimination of empty products. Biological replicates were sequenced for each condition.

Solexa libraries were subjected to cluster generation and 25 or 50 rounds of single-end sequencing using a unique sequencing primer (Churchman and Weissman, 2012). After processing and base calling by the Illumina Eland program, reads were first filtered by aligning with Bowtie 0.12.8 with default settings specifying --best against the reference noncoding RNA fasta file we created. This reference was created by taking the last 35nt of each sequence in BDGP5.68. Unmapped reads were then aligned against Dmel 5.3 as before. (Table S1)

MNase-seq & ChIP-seq

MNase-seq and ChIP-seq was performed as described (Weber et al., 2010) except without formaldehyde crosslinking and chromatin was solubilized with 80mM NaCl and needle extracted. Solexa libraries were constructed as previously described (Henikoff et al., 2011) with modifications. We used the TruSeq oligo design to enable barcoding of libraries with an AMPure XP to sample ratio of 1:1. Cluster generation and 25 rounds of paired-end sequencing by Illumina HiSeq 2000, was performed by the Fred Hutchinson Cancer Research Center Genomic and Shared Resource. Following processing and base calling by the Illumina Eland

program, reads with zero, one, or two, mismatches were mapped to Dmel 5.3 using Novoalign and default parameters where multiple hits were assigned to one location at random.

Data Analysis

The fraction of reads mapped at each nucleotide was multiplied by the total number of nucleotides mapped genome-wide to give a normalized count at that position. The 3'NT signal for each strand was normalized separately and was used at base-pair resolution (For 3'NT, each read was mapped to a single position, which corresponds to the 3' end of the nascent transcript). 3' exon end position was excluded from 3'NT analysis. MNase-seq and ChIP-seq reads were aggregated to 10 base-pair windows and only paired-end reads of length ≥76 and ≤160 bp were used. PRO-seq datasets were obtained from the GEO database, under the accession GSM1032758. The bedgraph file was parsed with each line as a read and normalized in identical manner as 3'NT data.

Enrichment over the flank mean was calculated as follows: the flank for each fly transcript (identifier: FBtr# in flybase, database version Dme1 5.3) was from TSS to TSS+2000 bp, unless interrupted by transcripts on same or opposite strand. The minimum flank considered was 500 bp. For plots of average 3'NT (or PRO-seq) relative to the nucleosome dyad, we define the flank as Dyad-120 to Dyad+120. For 3'NT or MNase-seq data, the mean enrichment was calculated for each flank. When plotting 3'NT or MNase-seq for a given transcript, the enrichment at a given position was divided by the mean enrichment of the flank to enable comparison between different transcripts. For most analyses, only those transcripts that had non-zero flank mean were used. Expression quartiles were determined based on the flank mean of transcripts.

MNase-seq enrichment between TSS-2000 and TSS+2000 was used for calling dyad positions. A 20-point running average and the first derivative was calculated around each position. The definition of a peak was a point at which the derivative was between -0.04 and +0.04; around which the derivative changed from negative to positive and where the MNase-seq signal was higher than the mean + 0.5*(standard error). The mean and standard error are calculated from ±10 positions around that point. If two peaks were within 100 bp, only the position with higher MNase-seq occupancy was retained as a peak. Comparison with DANPOS, an independent method (Chen et al., 2013a) reveals that a majority of nucleosome positions are similar (Figure 3-5)

To define stall positions, the mean and the standard deviation of the surrounding 200 bp is calculated (excluding stall positions). The position is defined as a stall if its 3'NT value is higher than the mean + 3*(standard deviation). Since the stall positions are to be excluded from the mean calculation at a given position, finding stalls is an iterative process: at each subsequent round of stall identification, stalls identified in the previous rounds are eliminated from mean calculation. This process is repeated until no new stalls are identified. The Stall Density is defined as fraction of transcripts having a stall at a given position relative to the nucleosome dyad. The Stall Fraction is defined as the fraction of stalls in a given region of a transcript, usually between dyad and dyad-100 bp.

Nucleosome occupancy and H2A.Z occupancy were calculated from MNase-seq and ChIP-seq data respectively, based on the called nucleosome dyad position. Occupancy is the sum of normalized reads between dyad-80 and dyad+80 for a given nucleosome of a given transcript. To calculate changes in H2A.Z occupancy upon knockdown, the H2A.Z occupancy for various nucleosomes was calculated for the H2A.Z and YL1 knockdown cells. We observed the +1, +2

and gene-body nucleosome occupancy to have significant linear correlation between knockdown and control (Figure 3-10), hence we transformed the knockdown occupancies based on the linear fits of nucleosome occupancy between control and knockdown. After the transformation, the ratio of H2A.Z occupancy between control and knockdown reflected the relative loss or gain of H2A.Z upon knockdown.

The 3'NT ratio of a given stretch of a transcript (between dyad-100 and dyad for a given nucleosome) is defined as the ratio of the sum of 3'NT reads in that stretch in the knockdown to the sum for the control. The 3'NT reads are normalized by the flank mean to enable comparison between different datasets. A similar ratio was also calculated for a window of 100 bp at defined positions from the TSS to generate the plot shown in Figure 3-11.

Accession Number

Genomic Data presented in this work are available from Gene Expression Omnibus under accession number GSE49106.

Chapter 4

Perspectives

Biochemical studies using purified components *in vitro* first showed that nucleosomes are strong barriers for both initiation and elongation of transcription (Kulaeva et al., 2013). Accordingly, cells evolved regulatory strategies to overcome this barrier and ensure precise gene transcription. Insight into how this happens *in vivo* first came from yeast genetics, which defined many of the factors involved (Rando and Winston, 2012). Yet, it wasn't until very recently, about the time that this dissertation research began, that we began to gain insight into where these factors act across the genome and the mechanisms involved. These insights were primarily fueled by technological advances in genomics, first microarrays and later sequencing, coupled with methodological advances such as chromatin immunoprecipitation (ChIP) and RNA sequencing (Park, 2009). These advances revolutionized molecular biology by empowering questions that were previously intractable. Suddenly it became possible to investigate the relationship between chromatin and transcription within cells at extremely high resolution and on a genome-wide scale.

Effect of transcription on nucleosomes

Investigating the totality of chromatin and transcription is a daunting challenge due to a vast number of components involved, each with potentially overlapping function. Thus, we focused on physical components, nucleosomes themselves and the enzyme that transcribes across nucleosomes, RNAPII. Nucleosomes are extensively modified *in viv*o, which might serve a regulatory role in transcription, however this has remained unclear. One of the most intriguing ways that nucleosomes are modified is through the incorporation of histone variants because they

are structurally distinct and are deposited outside of replication by dedicated machinery not utilized by canonical histones. We focused on histone variant H2A.Z because it is universal across all eukaryotes and is essential for viability in metazoans, thus any findings would have broad significance. The literature on H2A.Z has remained quite murky, however it was previously shown that H2A.Z is incorporated into the nucleosomes flanking the transcription start site (TSS) in yeast (Raisner et al., 2005; Zhang et al., 2005a). H2A.Z is deposited by the Swr1 chromatin remodeling complex (Mizuguchi et al., 2004), however it has remained unclear how H2A.Z is 'targeted' to a particular region of the genome or how this influences transcription.

The crystal structure of H2A.Z containing nucleosomes predicts that when H2A.Z is paired with H2A, a steric clash would result at loop1 (Suto et al., 2000). Yet, both homotypic (H2A.Z / H2A.Z) and heterotypic (H2A.Z / H2A) nucleosomes readily form *in vitro* and *in vivo* (Chakravarthy et al., 2004; Viens et al., 2006). Considering that these nucleosomes are structurally distinct, we hypothesized that they might have different properties, which could lead to different genome-wide localization. We found that homotypic H2A.Z nucleosomes were highly enriched over the bodies of active genes and showed evidence of disruption during transcription. This result suggested a new parsimonious model to explain H2A.Z enrichment, that H2A.Z is simply incorporated due to transcription-mediated eviction of H2A/H2B dimers and then replaced with H2A.Z. Later studies in mouse confirmed this finding, suggesting that homotypic H2A.Z is the predominant form over genes in other metazoans as well (Nekrasov et al., 2012). This parsimonious model relies only upon histone turnover, which is well established to be driven by transcription (Deal et al., 2010; Jamai et al., 2007; Kimura and Cook, 2001) and does not require any specific targeting, which to date, has not been conclusively shown to exist at

the majority of genes. In human and mouse cells, the orthologous Swr1 complexes have been shown to interact with transcription factors and are thought to be involved in 'targeting' the complex (Gallant-Behm et al., 2012; Gevry et al., 2007). However, this has only been shown for a few sites in the genome and likely reflects specific cases. Thus, our results support a general model for how H2A.Z becomes enriched at the vast majority of genes.

Recently, the INO80 complex was shown to conduct the reverse reaction of Swr1 by removal of H2A.Z and replacement with H2A (Papamichos-Chronakis et al., 2011). It is thus possible that the Drosophila INO80 complex might also remodel homotypic H2A.Z nucleosomes over genes. However, considering that we observe depletion of heterotypic nucleosomes, which would be an intermediate in the reaction, it seems unlikely that INO80 is a dominant player. Alternatively, a chaperone was recently identified in human cells, ANP32E, which specifically removes H2A.Z without requiring H2A replacement (Obri et al., 2014). In the absence of this chaperone, H2A.Z accumulates throughout the genome, especially at the +1 nucleosome, suggesting that ANP32E clears out H2A.Z making it more dynamic. While a homolog of this chaperone has not been identified in Drosophila, it is possible that a similar chaperone exists to facilitate H2A.Z dynamics. Considering that H2A.Z facilitates RNAPII loading (Hardy et al., 2009), ANP32E or INO80 may function to repress genes by keeping H2A.Z levels low. However, it is equally possible that they facilitate promoter clearance by removing H2A.Z and disrupting nucleosomes in front of RNAPII.

Effect of nucleosomes on transcription

Considering our finding that nucleosomes over active genes are homotypic H2A.Z, likely as a consequence of encounters with RNAPII, we next wanted to know what effect H2A.Z and

nucleosomes have on RNAPII transit. For many years a major open question has been how RNAPII transcribes across nucleosomal DNA *in vivo*, considering that nucleosomes are major barriers that cause backtracking and arrest *in vitro* (Bondarenko et al., 2006; Kireeva et al., 2005; Kireeva et al., 2002; Kulaeva et al., 2009). The limitation to solving this question has been the lack of a methodology that can map the position of RNAPII with sufficient resolution.

In order to overcome this limitation we developed an approach to comprehensively map the position of RNAPII at single nucleotide resolution. We took advantage of basic properties of the RNAPII complex, insolubility and extreme stability, to isolate nascent RNA and sequence the 3' end of nascent transcripts (3'NT), thus determining the position of the active site of the enzyme. This approach doesn't require the use of transgenes, tags, or immunopurification and captures all forms of RNAPII, including those that have encountered a barrier and backtracked, essentially obtaining a complete snapshot. Thus, this simple strategy can be easily adapted to any system and provides an advantage over low-resolution ChIP, run-on approaches that only capture elongation competent RNAPII (Core et al., 2008; Kwak et al., 2013), or NET-seq which requires solubilization (Churchman and Weissman, 2011), which is far from complete under native conditions (Kimura et al., 1999). We utilized this new technique coupled with high-resolution MNase-seq nucleosome maps to define where RNAPII stalls as it transcribes through nucleosomes and the magnitude.

We found that the first nucleosome (+1) was a much larger barrier than downstream positions that causes RNAPII to backtrack. The major consensus location where RNAPII stalls is just upon entry into all nucleosomes, just around a single helical turn of DNA. At nucleosomes further into the gene body we also observed stalling near the dyad axis, however this was not observed at the first two nucleosome positions. This suggests that transcription across

nucleosomes *in vivo* is context-dependent and very different from existing models *in vitro* and in yeast (Churchman and Weissman, 2011), where the consensus barrier is located near the dyad axis. Our findings are surprising because histone DNA contacts are strongest near the dyad axis but are relatively weak at the entry site (Hall et al., 2009) where the consensus RNAPII stall site is located. However, *in vitro* experiments typically assemble a canonical unmodified nucleosome on a template that strongly favors nucleosome formation, whereas nucleosomes *in vivo* are typically not found over their preferred sequence due to the action of chromatin remodelers. Additionally, in vitro experiments do not use a torsionally constrained template and must use high salt to promote transcription through nucleosomes, all of which could contribute to this discrepancy.

Interestingly, we find that H2A.Z reduced the magnitude of the nucleosomal barrier, which might explain why it is most enriched at the +1, where the barrier is largest. Our results suggest that H2A.Z reduces the nucleosome barrier by being easily evicted from nucleosomes. This might be due to a weaker interaction with the $(H3/H4)_2$ docking domain (Suto et al., 2000), increased dynamics by an unknown chaperone in Drosophila similar to ANP32E, or a chromatin remodeler such as INO80. Thus, H2A.Z might ease the refractory nature of nucleosomes both intrinsically and through other external factors. That H2A.Z can reduce the magnitude of the nucleosome barrier to RNAPII helps to explain why it is essential during development, where precise transcriptional regulation is highly important.

Here we have addressed two sides of H2A.Z biology, the effect of transcription on H2A.Z and the effect of H2A.Z on transcription, and general principles of how chromatin and transcription are related. Our results point to relatively parsimonious models to explain the deposition and function of H2A.Z and more generally, the role of chromatin in transcription, that

have remained enigmatic for quite some time. Overall, this work emphasizes the power of technological and methodological advances to expand our understanding of biology, which would not have been possible otherwise.

Chapter 5

Future Directions

Mechanism of H2A.Z in transcription

Our analysis showed that normalized H2A.Z occupancy is roughly anti-correlated with nucleosome turnover. This suggested that H2A.Z might reduce the magnitude of the nucleosome barrier to RNAPII by being easily evicted rather than necessitating complete eviction of the nucleosome. However, this analysis is correlation based and thus, it might also be possible that nucleosomes with lower turnover are more likely to retain H2A.Z. Thus, it would be informative to test causality by reducing H2A.Z levels and then measuring $(H3/H4)_2$ turnover using CATCH-IT (Deal et al., 2010). If our model is correct then we expect to see an increase in nucleosome turnover, reflective of having the more refractory H2A in place of H2A.Z, which necessitates complete nucleosome turnover.

A major focus of this work was on H2A.Z itself as a physical component of the nucleosome. However, H2A.Z is also post-translationally modified by acetylation, SUMOlation, ubiquitination, and methylation of lysines (Sevilla and Binda, 2014). Acetylation of H2A.Z is important for proper regulation of transcript levels in yeast (Halley et al., 2010) and correlates with gene expression in human cancer cells (Valdes-Mora et al., 2012). Considering that acetylation of histone tails destabilizes nucleosomes via charge neutralization (Brower-Toland et

al., 2005), it is possible that acetylation is a key contributor to H2A.Z's role in lessening the barrier in transcription. This could be easily tested by mutating the lysine residues in the tail of H2A.Z using the CrispR/cas9 genome-editing technique, which has been shown to be effective in Drosophila (Bassett and Liu, 2014). We could then determine the position and magnitude of RNAPII stalling using the 3'NT method to test whether stalling is increased.

Mechanism of RNAPII transit across nucleosomes

A major unanswered question from this work is why the +1 nucleosome is a much larger barrier to RNAPII. One possible explanation is that RNAPII requires a set of elongation factors in order to efficiently transcribe through any nucleosome. Recruitment of these factors may be tightly regulated, perhaps through modification of the C-terminus of RNAPII (Heidemann et al., 2013). Thus, it may not be that the +1 is special necessarily, but by being the first barrier, the +1 might block RNAPII until the necessary elongation factors are recruited. This might serve a redundant mechanism to ensure that RNAPII is transcribing in the right place and that the complex is properly loaded so that it doesn't become arrested downstream, which could result in genome instability (Nudler, 2012).

An alternative explanation relates to the topological state of DNA. As RNAPII transcribes along the DNA template, it induces positive supercoils in front and negative supercoils behind (Liu and Wang, 1987). Positive supercoiling is known to destabilize nucleosomes (Levchenko et al., 2005; Pfaffle et al., 1990; Teves and Henikoff, 2014) and differences in the topological state may determine the magnitude of the barrier. For example, the +1 nucleosome may be a larger barrier because less positive torsion has accumulated due to a shorter distance travelled by RNAPII. This model is supported by work in our lab showing that

upon Topo I inhibition, which primarily relaxes transcription-mediated torsion, the greatest increase in positive supercoiling occurred downstream of the +1 nucleosome. Additionally, downstream nucleosomes showed an increase in complete turnover after Topo I inhibition, whereas the +1 was less affected (Teves and Henikoff, 2014).

Future work could resolve these models, to determine why the +1 is larger. For the first model we could reduce nucleosome occupancy by knockdown of Asf1, which is an H3/H4 chaperone important for regulating genomic integrity and gene expression (Feser et al., 2010). We could then determine the composition of RNAPII complexes by mass spectrometry in order to determine which factors are not properly loaded. In order to test the second scenario, we could inhibit Topo I and use the 3'NT approach to measure the extent of RNAPII stalling through different positions. Here we expect to see that the +1 is relatively unaffected, whereas downstream positions would show a global increase in stalling, of similar magnitude to the +1 nucleosome.

An alternative direction utilizes a candidate gene knockdown approach coupled with 3'NT in order to investigate the general mechanism of transcription through nucleosomes. Candidates could be selected from screens in yeast that show strong genetic interactions with H2A.Z and are known to be involved in transcription elongation, which include PAF and FACT complexes (Santisteban et al., 2011). FACT is perhaps most promising as it eases transcription through nucleosomes *in vitro* and has been shown to colocalize and track with RNAPII *in vivo* (Orphanides et al., 1998; Saunders et al., 2003). Upon depletion of these complexes it would be interesting to determine how RNAPII transit across nucleosomes change, thus informing of the mechanism utilized by each factor during transcription across nucleosomes.

References

Abbott, D.W., Chadwick, B.P., Thambirajah, A.A., and Ausio, J. (2005). Beyond the Xi: macroH2A chromatin distribution and post-translational modification in an avian system. J Biol Chem *280*, 16437-16445.

Adam, M., Robert, F., Larochelle, M., and Gaudreau, L. (2001). H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. Mol Cell Biol *21*, 6270-6279.

Adam, S., Polo, S.E., and Almouzni, G. (2013). Transcription recovery after DNA damage requires chromatin priming by the H3.3 histone chaperone HIRA. Cell *155*, 94-106.
Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nature Reviews Genetics *13*, 720-731.

Adelman, K., Marr, M.T., Werner, J., Saunders, A., Ni, Z., Andrulis, E.D., and Lis, J.T. (2005). Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. Mol Cell *17*, 103-112.

Agelopoulos, M., and Thanos, D. (2006). Epigenetic determination of a cell-specific gene expression program by ATF-2 and the histone variant macroH2A. Embo J *25*, 4843-4853.
Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell *9*, 1191-1200.

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.

Allis, C.D., Glover, C.V., Bowen, J.K., and Gorovsky, M.A. (1980). Histone variants specific to the transcriptionally active, amitotically dividing macronucleus of the unicellular eucaryote, Tetrahymena thermophila. Cell *20*, 609-617.

Angelov, D., Molla, A., Perche, P.Y., Hans, F., Cote, J., Khochbin, S., Bouvet, P., and Dimitrov, S. (2003). The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. Mol Cell *11*, 1033-1041.

Banaszynski, L.A., Wen, D., Dewell, S., Whitcomb, S.J., Lin, M., Diaz, N., Elsasser, S.J., Chapgier, A., Goldberg, A.D., Canaani, E.*, et al.* (2013). Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. Cell *155*, 107-120.

Bancaud, A., Wagner, G., Conde, E.S.N., Lavelle, C., Wong, H., Mozziconacci, J., Barbi, M., Sivolob, A., Le Cam, E., Mouawad, L.*, et al.* (2007). Nucleosome chiral transition under positive torsional stress in single chromatin fibers. Mol Cell *27*, 135-147.

Bao, Y., Konesky, K., Park, Y.J., Rosu, S., Dyer, P.N., Rangasamy, D., Tremethick, D.J., Laybourn, P.J., and Luger, K. (2004). Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. Embo J *23*, 3314-3324.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bassett, A.R., and Liu, J.L. (2014). CRISPR/Cas9 and genome editing in Drosophila. J Genet Genomics *41*, 7-19.

Billon, P., and Cote, J. (2012). Precise deposition of histone H2A.Z in chromatin for genome expression and maintenance. Biochim Biophys Acta *1819*, 290-302.

Bintu, L., Ishibashi, T., Dangkulwanich, M., Wu, Y.Y., Lubkowska, L., Kashlev, M., and Bustamante, C. (2012). Nucleosomal elements that control the topography of the barrier to transcription. Cell *151*, 738-749.

Biswas, D., Dutta-Biswas, R., Mitra, D., Shibata, Y., Strahl, B.D., Formosa, T., and Stillman, D.J. (2006). Opposing roles for Set2 and yFACT in regulating TBP binding at promoters. Embo J *25*, 4479-4489.

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

Bonisch, C., and Hake, S.B. (2012). Histone H2A variants in nucleosomes and chromatin: more or less stable? Nucleic Acids Res *40*, 10719-10741.

Brower-Toland, B., Wacker, D.A., Fulbright, R.M., Lis, J.T., Kraus, W.L., and Wang, M.D. (2005). Specific contributions of histone tails and their acetylation to the mechanical stability of nucleosomes. J Mol Biol *346*, 135-146.

Bruce, K., Myers, F.A., Mantouvalou, E., Lefevre, P., Greaves, I., Bonifer, C., Tremethick, D.J., Thorne, A.W., and Crane-Robinson, C. (2005). The replacement histone H2A.Z in a hyperacetylated form is a feature of active genes in the chicken. Nucleic Acids Res *33*, 5633-5639.

Bryson, T.D., Weber, C.M., and Henikoff, S. (2010). Baculovirus-encoded protein expression for epigenomic profiling in Drosophila cells. Fly (Austin) *4*, 258-265.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell *56*, 549-561.

Buschbeck, M., Uribesalgo, I., Wibowo, I., Rue, P., Martin, D., Gutierrez, A., Morey, L., Guigo, R., Lopez-Schier, H., and Di Croce, L. (2009). The histone variant macroH2A is an epigenetic regulator of key developmental genes. Nat Struct Mol Biol *16*, 1074-1079.

Cai, H., and Luse, D.S. (1987). Transcription Initiation by Rna Polymerase-Ii Invitro - Properties of Preinitiation, Initiation, and Elongation Complexes. Journal of Biological Chemistry *262*, 298-304.

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

Chadwick, B.P., Valley, C.M., and Willard, H.F. (2001). Histone variant macroH2A contains two distinct macrochromatin domains capable of directing macroH2A to the inactive X chromosome. Nucleic Acids Res *29*, 2699-2705.

Chakravarthy, S., Bao, Y., Roberts, V.A., Tremethick, D., and Luger, K. (2004). Structural characterization of histone H2A variants. Cold Spring Harb Symp Quant Biol *69*, 227-234.

Chakravarthy, S., Gundimella, S.K., Caron, C., Perche, P.Y., Pehrson, J.R., Khochbin, S., and Luger, K. (2005). Structural characterization of the histone variant macroH2A. Mol Cell Biol *25*, 7616-7624.

Chakravarthy, S., and Luger, K. (2006). The histone variant macro-H2A preferentially forms "hybrid nucleosomes". J Biol Chem *281*, 25522-25531.

Chang, E.Y., Ferreira, H., Somers, J., Nusinow, D.A., Owen-Hughes, T., and Narlikar, G.J. (2008). MacroH2A allows ATP-dependent chromatin remodeling by SWI/SNF and ACF complexes but specifically reduces recruitment of SWI/SNF. Biochemistry *47*, 13726-13732.

Changolkar, L.N., Costanzi, C., Leu, N.A., Chen, D., McLaughlin, K.J., and Pehrson, J.R. (2007). Developmental changes in histone macroH2A1-mediated gene regulation. Mol Cell Biol *27*, 2758-2764.

Changolkar, L.N., Singh, G., and Pehrson, J.R. (2008). macroH2A1-dependent silencing of endogenous murine leukemia viruses. Mol Cell Biol *28*, 2059-2065.

Chen, K., Xi, Y., Pan, X., Li, Z., Kaestner, K., Tyler, J., Dent, S., He, X., and Li, W. (2013a). DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. Genome Res *23*, 341-351.

Chen, P., Zhao, J., Wang, Y., Wang, M., Long, H., Liang, D., Huang, L., Wen, Z., Li, W., Li, X., *et al.* (2013b). H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. Genes Dev *27*, 2109-2124.

Chen, Y., Chen, Q., McEachin, R., Cavalcoli, J., and Yu, X. (2014). H2A.B facilitates transcription elongation at methylated CpG loci. Genome Res.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368-373.

Churchman, L.S., and Weissman, J.S. (2012). Native elongating transcript sequencing (NET-seq). Curr Protoc Mol Biol *Chapter 4*, Unit 4 14 11-17.

Coleman-Derr, D., and Zilberman, D. (2012). Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. PLoS Genet *8*, e1002988.

Conerly, M.L., Teves, S.S., Diolaiti, D., Ulrich, M., Eisenman, R.N., and Henikoff, S. (2010). Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. Genome Res *20*, 1383-1390.

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.

Costanzi, C., and Pehrson, J.R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599-601.

Creyghton, M.P., Markoulaki, S., Levine, S.S., Hanna, J., Lodato, M.A., Sha, K., Young, R.A., Jaenisch, R., and Boyer, L.A. (2008). H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. Cell *135*, 649-661.

Csankovszki, G., Panning, B., Bates, B., Pehrson, J.R., and Jaenisch, R. (1999). Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. Nat Genet *22*, 323-324.

Cui, B., Liu, Y., and Gorovsky, M.A. (2006). Deposition and function of histone H3 variants in Tetrahymena thermophila. Mol Cell Biol *26*, 7719-7730.

Deal, R.B., Henikoff, J.G., and Henikoff, S. (2010). Genome-Wide Kinetics of Nucleosome Turnover Determined by Metabolic Labeling of Histones. Science *328*, 1161-1164.

Desai, N.A., and Shankar, V. (2003). Single-strand-specific nucleases. FEMS Microbiol Rev *26*, 457-491.

Dhillon, N., and Kamakaka, R.T. (2000). A histone variant, Htz1p, and a Sir1p-like protein, Esc2p, mediate silencing at HMR. Mol Cell *6*, 769-780.

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.

Doyen, C.M., An, W., Angelov, D., Bondarenko, V., Mietton, F., Studitsky, V.M., Hamiche, A., Roeder, R.G., Bouvet, P., and Dimitrov, S. (2006a). Mechanism of polymerase II transcription repression by the histone variant macroH2A. Mol Cell Biol *26*, 1156-1164.

Doyen, C.M., Montel, F., Gautier, T., Menoni, H., Claudet, C., Delacour-Larose, M., Angelov, D., Hamiche, A., Bednar, J., Faivre-Moskalenko, C.*, et al.* (2006b). Dissection of the unusual structural and functional properties of the variant H2A.Bbd nucleosome. Embo J *25*, 4234-4244.

Drane, P., Ouararhni, K., Depaux, A., Shuaib, M., and Hamiche, A. (2010). The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. Genes Dev *24*, 1253-1265.

Elsasser, S.J., Huang, H., Lewis, P.W., Chin, J.W., Allis, C.D., and Patel, D.J. (2012). DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition. Nature *491*, 560-565.

Faast, R., Thonglairoam, V., Schulz, T.C., Beall, J., Wells, J.R., Taylor, H., Matthaei, K., Rathjen, P.D., Tremethick, D.J., and Lyons, I. (2001). Histone variant H2A.Z is required for early mammalian development. Curr Biol *11*, 1183-1187.

Fernandez-Capetillo, O., Lee, A., Nussenzweig, M., and Nussenzweig, A. (2004). H2AX: the histone guardian of the genome. DNA Repair (Amst) *3*, 959-967.

Feser, J., Truong, D., Das, C., Carson, J.J., Kieft, J., Harkness, T., and Tyler, J.K. (2010). Elevated histone expression promotes life span extension. Mol Cell *39*, 724-735.

Filipescu, D., Szenker, E., and Almouzni, G. (2013). Developmental roles of histone H3 variants and their chaperones. Trends Genet *29*, 630-640.

Galburt, E.A., Grill, S.W., Wiedmann, A., Lubkowska, L., Choy, J., Nogales, E., Kashlev, M., and Bustamante, C. (2007). Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. Nature *446*, 820-823.

Gallant-Behm, C.L., Ramsey, M.R., Bensard, C.L., Nojek, I., Tran, J., Liu, M., Ellisen, L.W., and Espinosa, J.M. (2012). DeltaNp63alpha represses anti-proliferative genes via H2A.Z deposition. Genes Dev *26*, 2325-2336.

Gamble, M.J., Frizzell, K.M., Yang, C., Krishnakumar, R., and Kraus, W.L. (2010). The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing. Genes Dev *24*, 21-32.

Gautier, T., Abbott, D.W., Molla, A., Verdel, A., Ausio, J., and Dimitrov, S. (2004). Histone variant H2ABbd confers lower stability to the nucleosome. EMBO Rep *5*, 715-720.

Gehring, M., Bubb, K.L., and Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science *324*, 1447-1451.

Gevry, N., Chan, H.M., Laflamme, L., Livingston, D.M., and Gaudreau, L. (2007). p21 transcription is regulated by differential localization of histone H2A.Z. Genes Dev *21*, 1869-1881.

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

Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X.*, et al.* (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell *140*, 678-691.

Goldman, J.A., Garlick, J.D., and Kingston, R.E. (2010). Chromatin remodeling by imitation switch (ISWI) class ATP-dependent remodelers is stimulated by histone variant H2A.Z. J Biol Chem *285*, 4645-4651.

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.

Hall, M.A., Shundrovsky, A., Bai, L., Fulbright, R.M., Lis, J.T., and Wang, M.D. (2009). High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. Nat Struct Mol Biol *16*, 124-129.

Halley, J.E., Kaplan, T., Wang, A.Y., Kobor, M.S., and Rine, J. (2010). Roles for H2A.Z and its acetylation in GAL1 transcription and gene induction, but not GAL1-transcriptional memory. PLoS Biol *8*, e1000401.

Hamada, F.N., Park, P.J., Gordadze, P.R., and Kuroda, M.I. (2005). Global regulation of X chromosomal genes by the MSL complex in Drosophila melanogaster. Genes Dev *19*, 2289-2294.

Han, W., Li, X., and Fu, X. (2011). The macro domain protein family: structure, functions, and their potential therapeutic implications. Mutat Res *727*, 86-103.

Hardy, S., Jacques, P.E., Gevry, N., Forest, A., Fortin, M.E., Laflamme, L., Gaudreau, L., and Robert, F. (2009). The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition. PLoS Genet *5*, e1000687.

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

Heidemann, M., Hintermair, C., Voss, K., and Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochim Biophys Acta *1829*, 55-62.

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

Henikoff, S. (2009). Labile H3.3+H2A.Z nucleosomes mark 'nucleosome-free regions'. Nat Genet *41*, 865-866.

Henikoff, S., Henikoff, J.G., Sakai, A., Loeb, G.B., and Ahmad, K. (2009). Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome Res *19*, 460-469.

Hu, G., Cui, K., Northrup, D., Liu, C., Wang, C., Tang, Q., Ge, K., Levens, D., Crane-Robinson, C., and 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.

Huang, C., Zhang, Z., Xu, M., Li, Y., Li, Z., Ma, Y., Cai, T., and Zhu, B. (2013). H3.3-H4 tetramer splitting events feature cell-type specific enhancers. PLoS Genet *9*, e1003558.

Illingworth, R.S., Botting, C.H., Grimes, G.R., Bickmore, W.A., and Eskeland, R. (2012). PRC1 and PRC2 are not required for targeting of H2A.Z to developmental genes in embryonic stem cells. PLoS One *7*, e34848.

Ishibashi, T., Dryhurst, D., Rose, K.L., Shabanowitz, J., Hunt, D.F., and Ausio, J. (2009). Acetylation of vertebrate H2A.Z and its effect on the structure of the nucleosome. Biochemistry *48*, 5007-5017.

Ishibashi, T., Li, A., Eirin-Lopez, J.M., Zhao, M., Missiaen, K., Abbott, D.W., Meistrich, M., Hendzel, M.J., and Ausio, J. (2010). H2A.Bbd: an X-chromosome-encoded histone involved in mammalian spermiogenesis. Nucleic Acids Res *38*, 1780-1789.

Jamai, A., Imoberdorf, R.M., and Strubin, M. (2007). Continuous histone H2B and transcription-dependent histone H3 exchange in yeast cells outside of replication. Mol Cell *25*, 345-355.

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

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev *21*, 1519-1529.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.

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

Jullien, J., Astrand, C., Szenker, E., Garrett, N., Almouzni, G., and Gurdon, J.B. (2012). HIRA dependent H3.3 deposition is required for transcriptional reprogramming following nuclear transfer to Xenopus oocytes. Epigenetics Chromatin *5*.

Kelly, T.K., Miranda, T.B., Liang, G., Berman, B.P., Lin, J.C., Tanay, A., and Jones, P.A. (2010). H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. Mol Cell *39*, 901-911.

Kimura, H., and Cook, P.R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. J Cell Biol *153*, 1341-1353.

Kimura, H., Tao, Y., Roeder, R.G., and Cook, P.R. (1999). Quantitation of RNA polymerase II and its transcription factors in an HeLa cell: Little soluble holoenzyme but significant amounts of polymerases attached to the nuclear substructure. Mol Cell Biol *19*, 5383-5392.

Kireeva, M.L., Hancock, B., Cremona, G.H., Walter, W., Studitsky, V.M., and Kashlev, M. (2005). Nature of the nucleosomal barrier to RNA polymerase II. Mol Cell *18*, 97-108.

Kireeva, M.L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M., and Studitsky, V.M. (2002). Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. Mol Cell *9*, 541-552.

Kobor, M.S., Venkatasubrahmanyam, S., Meneghini, M.D., Gin, J.W., Jennings, J.L., Link, A.J., Madhani, H.D., and Rine, J. (2004). A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. PLoS Biol *2*, E131.

Komissarova, N., and Kashlev, M. (1997). RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. Journal of Biological Chemistry *272*, 15329-15338.

Kouzine, F., Gupta, A., Baranello, L., Wojtowicz, D., Ben-Aissa, K., Liu, J., Przytycka, T.M., and Levens, D. (2013). Transcription-dependent dynamic supercoiling is a short-range genomic force. Nat Struct Mol Biol *20*, 396-403.

Kouzine, F., Sanford, S., Elisha-Feil, Z., and Levens, D. (2008). The functional response of upstream DNA to dynamic supercoiling in vivo. Nat Struct Mol Biol *15*, 146-154.

Krogan, N.J., Keogh, M.C., Datta, N., Sawa, C., Ryan, O.W., Ding, H., Haw, R.A., Pootoolal, J., Tong, A., Canadien, V.*, et al.* (2003). A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. Mol Cell *12*, 1565-1576.

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

Kulaeva, O.I., Hsieh, F.K., Chang, H.W., Luse, D.S., and Studitsky, V.M. (2013). Mechanism of transcription through a nucleosome by RNA polymerase II. Biochim Biophys Acta *1829*, 76-83.

Kumar, S.V., and Wigge, P.A. (2010). H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. Cell *140*, 136-147.

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.

Lagha, M., Bothma, J.P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D.S., Zeitlinger, J.*, et al.* (2013). Paused Pol II Coordinates Tissue Morphogenesis in the Drosophila Embryo. Cell *153*, 976-987.

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-U215.

Levchenko, V., Jackson, B., and Jackson, V. (2005). Histone release during transcription: displacement of the two H2A-H2B dimers in the nucleosome is dependent on different levels of transcription-induced positive stress. Biochemistry *44*, 5357-5372.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707-719.

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

Li, G., Levitus, M., Bustamante, C., and Widom, J. (2005b). Rapid spontaneous accessibility of nucleosomal DNA. Nat Struct Mol Biol *12*, 46-53.

Li, W., Nagaraja, S., Delcuve, G.P., Hendzel, M.J., and Davie, J.R. (1993). Effects of histone acetylation, ubiquitination and variants on nucleosome stability. Biochem J *296 (Pt 3)*, 737-744.

Li, Z., Gadue, P., Chen, K., Jiao, Y., Tuteja, G., Schug, J., Li, W., and Kaestner, K.H. (2012). Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. Cell *151*, 1608-1616.

Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. Proc Natl Acad Sci U S A *84*, 7024-7027.

Liu, X., Li, B., and GorovskyMa (1996). Essential and nonessential histone H2A variants in Tetrahymena thermophila. Mol Cell Biol *16*, 4305-4311.

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.

Luk, E., Ranjan, A., Fitzgerald, P.C., Mizuguchi, G., Huang, Y., Wei, D., and Wu, C. (2010). Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome. Cell *143*, 725-736.

Luk, E., Vu, N.D., Patteson, K., Mizuguchi, G., Wu, W.H., Ranjan, A., Backus, J., Sen, S., Lewis, M., Bai, Y.*, et al.* (2007). Chz1, a nuclear chaperone for histone H2AZ. Mol Cell *25*, 357-368.

Madigan, J.P., Chotkowski, H.L., and Glaser, R.L. (2002). DNA double-strand break-induced phosphorylation of Drosophila histone variant H2Av helps prevent radiation-induced apoptosis. Nucleic Acids Res *30*, 3698-3705.

Malik, H.S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. Nat Struct Biol *10*, 882-891.

Marques, M., Laflamme, L., Gervais, A.L., and Gaudreau, L. (2010). Reconciling the positive and negative roles of histone H2A.Z in gene transcription. Epigenetics *5*, 267-272.

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

McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc Natl Acad Sci U S A *101*, 1525-1530.

Mendez, J., and Stillman, B. (2000). Chromatin association of human origin recognition complex, Cdc6, and minichromosome maintenance proteins during the cell cycle: Assembly of prereplication complexes in late mitosis. Mol Cell Biol *20*, 8602-8612.

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.

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

Mito, Y., Henikoff, J.G., and Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. Science *315*, 1408-1411.

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.

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.

Moses, A.M., Pollard, D.A., Nix, D.A., Iyer, V.N., Li, X.Y., Biggin, M.D., and Eisen, M.B. (2006). Large-scale turnover of functional transcription factor binding sites in Drosophila. PLoS Comput Biol *2*, e130.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat Genet *39*, 1507-1511.

Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327*, 335-338.

Nekrasov, M., Amrichova, J., Parker, B.J., Soboleva, T.A., Jack, C., Williams, R., Huttley, G.A., and Tremethick, D.J. (2012). Histone H2A.Z inheritance during the cell cycle and its impact on promoter organization and dynamics. Nat Struct Mol Biol *19*, 1076-1083.

Ng, R.K., and Gurdon, J.B. (2008). Epigenetic inheritance of cell differentiation status. Cell Cycle *7*, 1173-1177.

Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., Hihara, S., Frangakis, A.S., Imamoto, N., Ishikawa, T.*, et al.* (2012). Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. Embo J *31*, 1644-1653.

Nock, A., Ascano, J.M., Barrero, M.J., and Malik, S. (2012). Mediator-Regulated Transcription through the +1 Nucleosome. Mol Cell *48*, 837-848.

Nudler, E. (2012). RNA polymerase backtracking in gene regulation and genome instability. Cell *149*, 1438-1445.

Nudler, E., Mustaev, A., Lukhtanov, E., and Goldfarb, A. (1997). The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. Cell *89*, 33-41.

Nusinow, D.A., Sharp, J.A., Morris, A., Salas, S., Plath, K., and Panning, B. (2007). The histone domain of macroH2A1 contains several dispersed elements that are each sufficient to direct enrichment on the inactive X chromosome. J Mol Biol *371*, 11-18.

Obri, A., Ouararhni, K., Papin, C., Diebold, M.L., Padmanabhan, K., Marek, M., Stoll, I., Roy, L., Reilly, P.T., Mak, T.W.*, et al.* (2014). ANP32E is a histone chaperone that removes H2A.Z from chromatin. Nature *505*, 648-653.

Okuwaki, M., Kato, K., Shimahara, H., Tate, S., and Nagata, K. (2005). Assembly and disassembly of nucleosome core particles containing histone variants by human nucleosome assembly protein I. Mol Cell Biol *25*, 10639-10651.

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

Ouararhni, K., Hadj-Slimane, R., Ait-Si-Ali, S., Robin, P., Mietton, F., Harel-Bellan, A., Dimitrov, S., and Hamiche, A. (2006). The histone variant mH2A1.1 interferes with transcription by down-regulating PARP-1 enzymatic activity. Genes Dev *20*, 3324-3336.

Palmer, D., Snyder, L.A., and Blumenfeld, M. (1980). Drosophila nucleosomes contain an unusual histone-like protein. Proc Natl Acad Sci U S A *77*, 2671-2675.

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.

Park, P.J. (2009). ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet *10*, 669-680.

Park, Y.J., Dyer, P.N., Tremethick, D.J., and Luger, K. (2004). A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome. J Biol Chem *279*, 24274-24282.

Pchelintsev, N.A., McBryan, T., Rai, T.S., van Tuyn, J., Ray-Gallet, D., Almouzni, G., and Adams, P.D. (2013). Placing the HIRA histone chaperone complex in the chromatin landscape. Cell Rep *3*, 1012-1019.

Pfaffle, P., Gerlach, V., Bunzel, L., and Jackson, V. (1990). In vitro evidence that transcription-induced stress causes nucleosome dissolution and regeneration. J Biol Chem *265*, 16830-16840.

Pina, B., and Suau, P. (1987). Changes in histones H2A and H3 variant composition in differentiating and mature rat brain cortical neurons. Dev Biol *123*, 51-58.

Placek, B.J., Huang, J., Kent, J.R., Dorsey, J., Rice, L., Fraser, N.W., and Berger, S.L. (2009). The Histone Variant H3.3 Regulates Gene Expression during Lytic Infection with Herpes Simplex Virus Type 1. J Virol *83*, 1416-1421.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432-445.

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

Rando, O.J., and Winston, F. (2012). Chromatin and Transcription in Yeast. Genetics *190*, 351-387.

Ranjan, A., Mizuguchi, G., FitzGerald, P.C., Wei, D., Wang, F., Huang, Y., Luk, E., Woodcock, C.L., and Wu, C. (2013). Nucleosome-free region dominates histone acetylation in targeting SWR1 to promoters for H2A.Z replacement. Cell *154*, 1232-1245.

Ratnakumar, K., Duarte, L.F., LeRoy, G., Hasson, D., Smeets, D., Vardabasso, C., Bonisch, C., Zeng, T., Xiang, B., Zhang, D.Y.*, et al.* (2012). ATRX-mediated chromatin association of histone variant macroH2A1 regulates alpha-globin expression. Genes Dev *26*, 433-438.

Ray-Gallet, D., Woolfe, A., Vassias, I., Pellentz, C., Lacoste, N., Puri, A., Schultz, D.C., Pchelintsev, N.A., Adams, P.D., Jansen, L.E.*, et al.* (2011). Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. Mol Cell *44*, 928-941.

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

Ridgway, P., Brown, K.D., Rangasamy, D., Svensson, U., and Tremethick, D.J. (2004). Unique residues on the H2A.Z containing nucleosome surface are important for Xenopus laevis development. J Biol Chem *279*, 43815-43820.

Rocha, E., Davie, J.R., van Holde, K.E., and Weintraub, H. (1984). Differential salt fractionation of active and inactive genomic domains in chicken erythrocyte. J Biol Chem *259*, 8558-8563.

Rougvie, A.E., and Lis, J.T. (1990). Postinitiation transcriptional control in Drosophila melanogaster. Mol Cell Biol *10*, 6041-6045.

Sakai, A., Schwartz, B.E., Goldstein, S., and Ahmad, K. (2009). Transcriptional and Developmental Functions of the H3.3 Histone Variant in Drosophila. Current Biology *19*, 1816-1820.

Samkurashvili, I., and Luse, D.S. (1996). Translocation and transcriptional arrest during transcript elongation by RNA polymerase II. Journal of Biological Chemistry *271*, 23495-23505.

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.

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.

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.

Saunders, A., Werner, J., Andrulis, E.D., Nakayama, T., Hirose, S., Reinberg, D., and Lis, J.T. (2003). Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. Science *301*, 1094-1096.

Sawatsubashi, S., Murata, T., Lim, J., Fujiki, R., Ito, S., Suzuki, E., Tanabe, M., Zhao, Y., Kimura, S., Fujiyama, S.*, et al.* (2010). A histone chaperone, DEK, transcriptionally coactivates a nuclear receptor. Genes Dev *24*, 159-170.

Schneiderman, J.I., Orsi, G.A., Hughes, K.T., Loppin, B., and Ahmad, K. (2012). Nucleosome-depleted chromatin gaps recruit assembly factors for the H3.3 histone variant. Proc Natl Acad Sci U S A *109*, 19721-19726.

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

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat Genet *38*, 700-705.

Selby, C.P., Drapkin, R., Reinberg, D., and Sancar, A. (1997). RNA polymerase II stalled at a thymine dimer: footprint and effect on excision repair. Nucleic Acids Res *25*, 787-793.

Selth, L.A., Sigurdsson, S., and Svejstrup, J.Q. (2010). Transcript Elongation by RNA Polymerase II. Annu Rev Biochem *79*, 271-293.

Sevilla, A., and Binda, O. (2014). Post-translational modifications of the histone variant H2AZ. Stem Cell Res *12*, 289-295.

Sheinin, M.Y., Li, M., Soltani, M., Luger, K., and Wang, M.D. (2013). Torque modulates nucleosome stability and facilitates H2A/H2B dimer loss. Nat Commun *4*, 2579.

Sigurdsson, S., Dirac-Svejstrup, A.B., and Svejstrup, J.Q. (2010). Evidence that Transcript Cleavage Is Essential for RNA Polymerase II Transcription and Cell Viability. Mol Cell *38*, 202-210.

Sims, R.J., 3rd, Millhouse, S., Chen, C.F., Lewis, B.A., Erdjument-Bromage, H., Tempst, P., Manley, J.L., and Reinberg, D. (2007). Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. Mol Cell *28*, 665-676.

Soboleva, T.A., Nekrasov, M., Pahwa, A., Williams, R., Huttley, G.A., and Tremethick, D.J. (2012). A unique H2A histone variant occupies the transcriptional start site of active genes. Nat Struct Mol Biol *19*, 25-30.

Suto, R.K., Clarkson, M.J., Tremethick, D.J., and Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nat Struct Biol *7*, 1121-1124.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell *116*, 51-61.

Talbert, P.B., and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol *11*, 264-275.

Tamura, T., Smith, M., Kanno, T., Dasenbrock, H., Nishiyama, A., and Ozato, K. (2009). Inducible Deposition of the Histone Variant H3.3 in Interferon-stimulated Genes. Journal of Biological Chemistry *284*, 12217-12225.

Teves, S.S., and Henikoff, S. (2011). Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. Genes Dev *25*, 2387-2397.

Teves, S.S., and Henikoff, S. (2014). Transcription-generated torsional stress destabilizes nucleosomes. Nat Struct Mol Biol *21*, 88-94.

Thakar, A., Gupta, P., Ishibashi, T., Finn, R., Silva-Moreno, B., Uchiyama, S., Fukui, K., Tomschik, M., Ausio, J., and Zlatanova, J. (2009). H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. Biochemistry *48*, 10852-10857.

Thakar, A., Gupta, P., McAllister, W.T., and Zlatanova, J. (2010). Histone variant H2A.Z inhibits transcription in reconstituted nucleosomes. Biochemistry *49*, 4018-4026.

Thambirajah, A.A., Dryhurst, D., Ishibashi, T., Li, A., Maffey, A.H., and Ausio, J. (2006). H2A.Z stabilizes chromatin in a way that is dependent on core histone acetylation. J Biol Chem *281*, 20036-20044.

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

Tochio, N., Umehara, T., Munemasa, Y., Suzuki, T., Sato, S., Tsuda, K., Koshiba, S., Kigawa, T., Nagai, R., and Yokoyama, S. (2010). Solution structure of histone chaperone ANP32B: interaction with core histones H3-H4 through its acidic concave domain. J Mol Biol *401*, 97-114.

Tolstorukov, M.Y., Goldman, J.A., Gilbert, C., Ogryzko, V., Kingston, R.E., and Park, P.J. (2012). Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. Mol Cell *47*, 596-607.

Valdes-Mora, F., Song, J.Z., Statham, A.L., Strbenac, D., Robinson, M.D., Nair, S.S., Patterson, K.I., Tremethick, D.J., Stirzaker, C., and Clark, S.J. (2012). Acetylation of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. Genome Res *22*, 307-321.

van Daal, A., and Elgin, S.C. (1992). A histone variant, H2AvD, is essential in Drosophila melanogaster. Mol Biol Cell *3*, 593-602.

van Holde, K.E., Lohr, D.E., and Robert, C. (1992). What happens to nucleosomes during transcription? J Biol Chem *267*, 2837-2840.

Viens, A., Mechold, U., Brouillard, F., Gilbert, C., Leclerc, P., and Ogryzko, V. (2006). Analysis of human histone H2AZ deposition in vivo argues against its direct role in epigenetic templating mechanisms. Mol Cell Biol *26*, 5325-5335.

Wang, Z., and Droge, P. (1996). Differential control of transcription-induced and overall DNA supercoiling by eukaryotic topoisomerases in vitro. Embo J *15*, 581-589.

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.

Weber, C.M., Henikoff, J.G., and Henikoff, S. (2010). H2A.Z nucleosomes enriched over active genes are homotypic. Nat Struct Mol Biol *17*, 1500-1507.

Weber, C.M., Ramachandran, S., and Henikoff, S. (2014). Nucleosomes are context-specific, H2A.Z modulated barriers to RNA polymerase Mol Cell.

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

Wu, R.S., Tsai, S., and Bonner, W.M. (1982). Patterns of histone variant synthesis can distinguish G0 from G1 cells. Cell *31*, 367-374.

Wu, W.H., Alami, S., Luk, E., Wu, C.H., Sen, S., Mizuguchi, G., Wei, D., and Wu, C. (2005). Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. Nat Struct Mol Biol *12*, 1064-1071.

Wuarin, J., and Schibler, U. (1994). Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. Mol Cell Biol *14*, 7219-7225.

Wysocka, J., Reilly, P.T., and Herr, W. (2001). Loss of HCF-1-chromatin association precedes temperature-induced growth arrest of tsBN67 cells. Mol Cell Biol *21*, 3820-3829.

Xie, W., Song, C., Young, N.L., Sperling, A.S., Xu, F., Sridharan, R., Conway, A.E., Garcia, B.A., Plath, K., Clark, A.T.*, et al.* (2009). Histone h3 lysine 56 acetylation is linked to the core transcriptional network in human embryonic stem cells. Mol Cell *33*, 417-427.

Xu, F., Zhang, K., and Grunstein, M. (2005). Acetylation in histone H3 globular domain regulates gene expression in yeast. Cell *121*, 375-385.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033-1037.

Yamaguchi, Y., Shibata, H., and Handa, H. (2013). Transcription elongation factors DSIF and NELF: Promoter-proximal pausing and beyond. Bba-Gene Regul Mech *1829*, 98-104.

Yang, J.H., Song, Y., Seol, J.H., Park, J.Y., Yang, Y.J., Han, J.W., Youn, H.D., and Cho, E.J. (2011). Myogenic transcriptional activation of MyoD mediated by replication-independent histone deposition. Proc Natl Acad Sci U S A *108*, 85-90.

Yen, K., Vinayachandran, V., and Pugh, B.F. (2013). SWR-C and INO80 chromatin remodelers recognize nucleosome-free regions near +1 nucleosomes. Cell *154*, 1246-1256.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916-919.

Zentner, G.E., and Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. Nat Struct Mol Biol *20*, 259-266.

Zhang, H., Roberts, D.N., and Cairns, B.R. (2005a). Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. Cell *123*, 219-231.

Zhang, R., Poustovoitov, M.V., Ye, X., Santos, H.A., Chen, W., Daganzo, S.M., Erzberger, J.P., Serebriiskii, I.G., Canutescu, A.A., Dunbrack, R.L.*, et al.* (2005b). Formation of MacroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA. Dev Cell *8*, 19-30.

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

Zhou, J., Fan, J.Y., Rangasamy, D., and Tremethick, D.J. (2007). The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. Nat Struct Mol Biol *14*, 1070-1076.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature *456*, 125-129.
Zlatanova, J., and Thakar, A. (2008). H2A.Z: view from the top. Structure *16*, 166-179.