Histone H3 Lysine 36 modification distinguishes transcribed and non-transcribed regions of the *S. cerevisiae* genome

Bhargavi Rao

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Genetics and Molecular Biology.

Chapel Hill
2007

Approved by
Advisor: Jason Lieb
Advisor: Brian Strahl
Reader: Charles Perou
Reader: Tom Petes
Reader: Jeff Sekelsky

ABSTRACT

Bhargavi Rao: Histone H3 Lysine 36 modification distinguishes regulatory and coding functions of the *S. cerevisiae* genome

(Under the direction of Drs. Jason Lieb and Brian Strahl)

Chromatin plays a dynamic role in regulating gene transcription. Regulation of accessibility of DNA template is mediated in part by nucleosome occupancy such that nucleosomes are relatively depleted upstream of genes and relatively enriched in the coding regions. One of the factors that influence this differential nucleosome occupancy is histone post translational modifications. One such modification is dimethylation of histone H3 at Lysine 36 (H3K36me2). It is mediated by Set2, a histone methyl transferase (HMT) in yeast which had been shown to associate with RNA polymerase II (RNA pol-II) during transcription elongation at individual loci. To study the role of Set2 in gene regulation, I sought to determine the genome wide localization of H3K36me2.

Using chromatin immunoprecipitation followed by DNA microarray hybridization (ChIP-chip), we show that H3K36me2 is predominantly localized to RNA pol-II transcribed regions and is depleted in the regulatory (promoter) regions genome-wide. Mating loci, telomeres, RNA pol-III transcribed regions have scarce or low levels of H3K36me2. H3K36me2 modification begins within RNA pol-II transcribed ORFs at approximately same location, independent of the length of the ORF. This further confirms that Set2 associates with RNA pol-II after the initiation phase of transcription. Levels of H3K36me2 do not correlate with the transcriptional frequencies of genes. However, genes that are transcribed at some detectably level tend to have higher levels of H3K36me2 than genes that are completely repressed.

H3K36me2 therefore acts as a mark that demarcates coding and regulatory regions. The function of such a mark became clear with the finding by other groups that localization of Set2 and H3K36me2 at coding regions was essential for maintaining the fidelity of transcriptional initiation. Absence of Set2 leads to hyperacetylation in the coding regions and, as a consequence, aberrant initiation events.

My studies show that H3K36me2 is a chromatin mark that demarcates functionally distinct regions of the genome by marking the coding regions specifically. Studies by others show that this localization of H3K36me2 is important for maintaining proper chromatin structure. H3 Lysine 36 is also acetylated and ChIP-chip analysis showed that H3K36ac is enriched in the promoter regions in the entire yeast genome. The function of H3K36ac is not yet known but it is possible that one way H3K36me2 is restricted to the coding regions by acetylating this residue in the regulatory regions. Another way organisms demarcate specific functional boundaries is by restricting tri methyl Lysine 4 at histone H3 (H3K4me3) to the 5′ end of coding regions. Ctk1, a kinase that has been shown to phosphorylate Serine 2 of C-terminal domain (CTD) of RNA pol-II was shown to regulate the levels of H3K4me3. Ctk1 is required for the recruitment of Set2 to RNA pol-II. My genome wide studies show that absence of Ctk1 causes spreading of H3K4me3 into the 3′ region of ORFs globally resulting in disruption of chromatin structure within the ORFs and occurrence of aberrant transcription initiation. These studies show that specific histone modification patterns are important for maintaining chromatin structure. Organisms have developed multiple mechanisms to ensure proper localization of these modifications disruption of which could cause disturbances in transcriptional programs.

DEDICATION

I would like to dedicate this thesis to my parents, Sridhar and Leela Rao and my husband, Hemanth Bhat who have always been there to support me through the roughest times.

Dad, this is for you. I miss you.

ACKNOWLEDGEMENTS

I would like to thank my advisors, Dr. Jason Lieb and Dr. Brian Strahl for being the best advisors I could have ever asked for. It has been a rewarding experience to work under their guidance. I will forever be indebted to them for being patient, understanding, and inspiring.

I would also like to thank members of the Lieb and Strahl labs, Michael Buck, Sean Hanlon, Asia Mieczkowska, Christina Whittle, Kelby Kizer, Tiaojiang Xiao, Yoichiro Shibata, Stephanie Morris, Nick Laribee, Sevinc Ercan, Paul Giresi, and Andrea Gossett.

And a special thanks to the past members of Strahl and Lieb labs, Hana Hall, Ostranda Williams and Cheol-Koo Lee.

Thank You to all of you. It has been a privilege working with you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION

Chromatin Structure

In eukaryotic organisms, genomic DNA is associated with proteins called histones to form chromatin. The fundamental unit of chromatin, the nucleosome, contains ~146 bp of DNA wrapped around a core consisting of two copies each of the four histone proteins, H2A, H2B, H3 and H4. The nucleosome was first described by Roger Kornberg in 1974 as the fundamental repeating unit of the chromosome, which appeared as a "beads on a string" structure (Kornberg and Thomas, 1974). Nucleosome cores are separated by a variable length of linker DNA that is associated with a single molecule of a fifth histone called the linker histone, H1. Nucleosomes not only provide the highest level of compaction to fit meters of double stranded DNA into the nucleus but they are important in the regulation of transcription by preventing RNA polymerase from unnecessarily accessing the regulatory regions of genes.

How does transcription affect nucleosome distribution?

Nucleosomes form a barrier to the passage of RNA polymerase II (RNA pol-II). For polymerase to traverse the nucleosomes requires removal of histone H2A/H2B (Kireeva et al., 2005; Kireeva et al., 2002; Studitsky et al., 2004). Genome wide studies have shown that nucleosomes are lost upon increased transcription (Lee et al., 2004). Nucleosomes that are disassembled in front of the transcribing polymerase associate with elongation factors, Spt6p and Spt16p. Spt16p is a subunit of the FACT (facilitates chromatin transcription) complex. FACT has been shown to interact

with histone H2A/H2B dimer and enhances transcription through nucleosomes in vivo suggesting that the complex may be promoting nucleosome disassembly during transcription (Orphanides et al., 1999). Spt6p on the other hand interacts with histones H3/H4 (Belotserkovskaya and Reinberg, 2004). These factors associate with histones and are necessary for restoring the chromatin structure after the polymerase has passed by. This is evident by the fact that in the absence of these factors genes show lower nucleosome density and exhibit transcription initiation from within the coding regions (Kaplan et al., 2003; Mason and Struhl, 2003).

Heterogeneous Nucleosome Distribution

Early studies on chromatin started with the discovery of DNase I hypersensitive sites at the promoter regions of active genes (Weintraub and Gourdine; 1976, Elgin SC; 1981). Subsequently it was suggested that chromatin had an active role in gene regulation. These studies showed that nucleosomes caused a hindrance for transcription factor accessibility to DNA and that gene activation involved removal or alteration of the nucleosomes (Knezetic and Luse, 1986; Levy and Noll, 1981; Lorch et al., 1987; Workman and Roeder, 1987).

Although the role of chromatin in gene regulation had been hypothesized a long time ago it is only recently that extensive studies on the nucleosome distribution and occupancy have clearly demonstrated their role in gene activity. The nucleosome position regulates the accessibility of the underlying DNA sequence to enable transcription. Chromatin immunoprecipitation studies in *Saccharomyces cerevisiae* have shown occurrence of well positioned nucleosomes genome-wide with regions of nucleosome free regions. On one of the most characterized and well studied genes, *PHO5*, it was shown that upon induction, the nucleosomes at the promoter are

hyperacetylated and subsequently lost (Reinke and Horz, 2003). Other studies suggested that rather than being lost completely the nucleosomes are unfolded and there possibly exists an equilibrium between the removal of the nucleosomes and their reformation .Whole genome microarray studies have shown that nucleosomes are depleted from active regulatory regions (Bernstein et al., 2004; Lee et al., 2004) and the level of nucleosome occupancy is inversely proportional to transcription initiation rate at the promoter. Rando and colleagues have used a tiled microarray approach to identify nucleosome occupancy at high resolution over 482 kilobases of *Saccharomyces cerevisiae* DNA, including most of Chromosome III (Yuan et al., 2005). This study mapped 2278 nucleosomes and found that most of them are well positioned. This study also showed that at the promoters of RNA pol-II transcribed genes about ~200bp upstream of the start codon there was a nucleosome region flanked by two well-positioned nucleosomes.

How is this nucleosome heterogeneity established and maintained?

This differential distribution of nucleosomes is critical for the passage of RNA pol-II during transcription elongation. There are several factors that influence nucleosome distribution including transcription factors, DNA sequence, ATP-dependent nucleosome remodeling complexes, histone variants and histone modifications.

Transcription factors

Although it was known that the *PHO5* promoter becomes nuclease hypersensitive upon transcriptional activation (Bergman and Kramer, 1983), recent evidence has shown that the Pho4 transcription factor binds to the promoter of *PHO5* prior to nucleosome disassembly and that this binding requires acetylation of nucleosomes by NuA4 complex (Adkins and Tyler, 2004) (Nourani et al., 2004). Another

transcription factor, Rap1p was shown to enable Gcn4p binding for the activation of *HIS4* promoter (Yu and Morse, 1999). Reb1p, a transcription factor can also generate a nucleosome-free region flanked by H2A.Z-containing nucleosomes (Raisner et al., 2005) .These studies indicate that transcription factor binding to the promoter DNA prevents nucleosome binding.

DNA sequence

Another factor that affects nucleosome distribution is the underlying DNA sequence. Two recent, and exciting studies have shown that the DNA sequence not only codes the genomic information but also determines how it is packaged into nucleosomes. Widom and colleagues use a combination of experimental and computational methods to study the sequence of DNA preferentially associated with nucleosomes (Segal et al., 2006). The study involved isolating regions of DNA that stably associated with nucleosomes and used these sequences to construct a probabilistic model that represents the positioning of nucleosomes. Pugh and colleagues used the frequency of AA and TT along DNA to define a 'nucleosome positioning sequence'(NPS) (Ioshikhes et al., 2006). Both these groups compared their models to the *in vivo* nucleosome positioning data and show that underlying DNA sequence is a very good predictor of nucleosome position. The latter study also predicted that the regions that have a fairly compact nucleosome positioning sequence are the ones that have a TATA box and are regulated by chromatin modifying and remodeling factors.

Histone variants

Nucleosomes containing the histone variant H2A.Z flank the nucleosome-free regions of inactive yeast promoters, and are preferentially displaced during gene activation (Guillemette et al., 2005; Li et al., 2005; Raisner et al., 2005; Zhang et al., 2005). H2A.Z/H2B dimers are less stable than H2A/H2B (Placek et al., 2005) and therefore H2A.Z-containing nucleosomes are thought to be more easily

displaced from chromatin than H2A-containing nucleosomes (Zhang et al., 2005). Although it is not very clear if there is a correlation between the transcriptional frequency of a gene and the presence of H2A.Z, it is likely that presence of H2A.Z poises a gene for rapid induction since there is a lag of several minutes in the heat-shock induction of *YDC1* in *htz1* mutant cells (Zhang et al., 2005).

Histone Modifications

Histone modifications also contribute to the regulation of nucleosome stability. The N- and C-terminal tail domains of histones are rich in arginine and lysine residues and are available for protein-DNA and protein-protein interactions. These tails are subject to post-translational modifications, including acetylation, phosphorylation, ubiquitination and methylation (Grant et al; 2001; van Holde; 1989). With the development of antibodies that recognize each histone modification specifically and their usage in chromatin immunoprecipitation followed by DNA microarrays (ChIP-chip), there have been advances in the genome-wide mapping of these modifications.

Histone tail modifications affect chromatin structure and its functional properties like transcription, cell division, repair and other biological processes. The singular as well as combinatorial modifications could affect chromatin function through distinct mechanisms: Modifications could directly interfere with the integrity and stability of a single nucleosome or an array of nucleosomes. Bulk acetylation, for example, has been shown to alter the secondary structure of the histone tail and weaken histone-DNA interactions resulting in the opening of the chromatin structure to allow transcription (Hong et al., 1993).. Modifications can also reduce internucleosomal interactions and chromatin folding caused by changes in the net charge of the histone tails (Luger et al., 1997). The modifications could also affect the binding of regulatory

factors (Kwon et al., 2000; Lee et al., 1993; Nightingale et al., 1998; Steger et al., 1998; Vettese-Dadey et al., 1996). HP1/Swi6p is a chromodomain containing protein involved in the formation of heterochromatin and in the silencing of gene expression by binding methylated H3 Lys-9 (H3K9) (Jenuwein and Allis, 2001; Zhang and Reinberg, 2001). Conversely, certain histone modification patterns appear to prevent the binding of chromatin associated mediators or effector modules. For example, methylation at Lys-4 and not Lys-9 prevents binding of transcriptional repressor complex NuRD (nucleosome remodeling and deacetylase), thus disrupting the association of histones with a repressor complex (Zegerman et al., 2002). More recently studies from several groups have shown that the plant homeodomain (PHD) finger specifically binds methylated lysine and regulates gene expression (Li et al., 2006; Pena et al., 2006; Shi et al., 2006; Wysocka et al., 2006).

Many of the enzymes that post-translationally modify histones display a high degree of specificity not only towards a particular amino acid residue, but also towards the pre-existing modification state of their histone substrates. *In vivo* and *in vitro* data suggest that acetylation of H3K14 and phosphorylation of H3 at serine-10 (H3S10) are coupled and that H3 acetylation by yeast Gcn5p is enhanced by phosphorylation at H3S10 (Cheung et al., 2000; Clements et al., 2003; Lo et al., 2000). Additionally, phosphorylation of H3S10 prevents binding of HP1 to methylated H3K9 (Fischle et al., 2005). Chemical modification of a nearby residue has also been shown to affect the substrate recognition ability of the histone modifying enzyme. Isomerization of proline-38 prevents methylation of H3K36 by Set2p (Nelson et al., 2006).

The large number of modifications, along with the fact that functional interplay exists between them, has led to the suggestion of a ‘histone code’ that controls chromatin biology, most notably gene transcription (Jenuwein and Allis, 2001; Strahl and Allis, 2000; Turner, 2000). In other

words, a combination of specific sets of modifications (code) at any region may define its transcriptional state as it is 'read' by a set of proteins which translate the code into an 'active or repressed state' of chromatin.

Histone methylation

Histone methylation occurs on both arginine and lysine residues. While arginine residues can be mono- or dimethylated in symmetric or asymmetric configurations, the epsilon-amino group of lysine residues can be mono-, di- or trimethylated (van Holde; 1989; von Holt; 1989). Histone lysine methylation is known to occur on histone H3 (lysines 4, 9, 27, 36 and 79) and histone H4 (lysine 20) (Strahl et al., 1999; van holde, 1988)**.** Lysine residues are methylated by a class of Histone methyl transferases (HMTs) most of which contain the SET domain which took is name from the *Drosophila* genes *Su(var) 3-9, Enhancer of zeste* and *Trithorax* (Jenuwein et al., 1998). The first SET domain containing histone lysine methyltransferase reported was the mammalian Suv39h1 and its fission yeast homolog Clr4p, which methylate histone H3K9 (Rea et al., 2000). The residues that are methylated in *S.cerevisae* are H3K4, H3K36 and H3K79 in the globular domain of histone H3. The enzymes responsible are Set1p, Set2p and Dot1p respectively.

Lys-4 methylation and transcriptional regulation

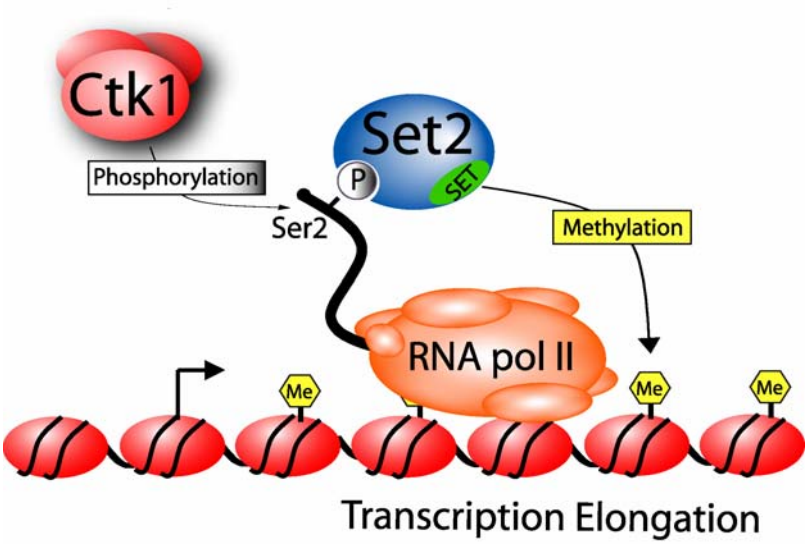
Most extensively studied histone methylation is Lys-4 methylation of histone H3. In *S.cerevisiae*, Set1p methylates H3K4 (Briggs et al., 2001). Set1p targeting to DNA coding regions is mediated by elongating RNA pol-II and requires phosphorylation at Ser-5 of CTD (Ng et al., 2003b). Ser-5 of CTD is phosphorylated during initiation and early elongation stages of transcription. Phosphorylated Ser-2 is specific to later stages of elongation. Since H3K4me3 and Set1p has been shown to be localized specifically to 5′ regions of actively transcribing genes

(Santos-Rosa et al., 2002), it is speculated that Set1p associates with RNA pol-II early in the elongation phase. H3K4me2, on the other hand, was found to be present mostly in the middle of the genes and H3K4me1 was enriched at the 3′ end of the genes (Pokholok et al., 2005). Set1p has also been shown to play a Sir2p-independent role in rDNA transcriptional silencing and this role involves methylation of histone H3 (Bryk et al., 2002). The *SET1* null mutant displays a decrease in transcriptional activity on a genome-wide scale, indicating that Set1p positively influences transcription by histone H3K4me (Bernstein et al., 2002; Boa et al., 2003).

Lys- 36 methylation and transcriptional regulation

Histone H3 Lys-36 methylation in *S.cerevisiae* is mediated by the methyltransferase Set2p (Strahl et al., 2002). In recent studies, Set2p has been shown to interact with RNA pol-II via its Set2 Rpb1 interacting (SRI) domain (Kizer et al., 2005; Krogan et al., 2003; Li et al., 2002; Xiao et al., 2003). This interaction requires the CTD of the largest subunit of RNA pol-II. Specifically, Set2p preferentially associates with Ser-2 phosphorylated form of RNA pol-II. Since Ser-5 of the CTD is phosphorylated during initiation of transcription and Ser-2 is phosphorylated during elongation, the physical association of Set2p with this specific modification suggests a role for this enzyme in transcription elongation mediated by RNA pol-II. Deletion of *CTK1*, the gene encoding a cyclin-dependent kinase that phosphorylates Ser-2 of the CTD, abolishes H3K36me (Xiao et al., 2003). All the subunits of the elongation factor, PAF complex (Paf1p, Rtf1p, Cdc73p, Ctr9p and Leo1p), associate with RNA pol-II and have shown to be localized to coding regions. When genes encoding Cdc73p and Rtf1p are deleted, a significant reduction in Set2p's recruitment to the *PMA1* gene and loss of H3K36me is found (Krogan et al., 2003), thus further

suggesting a role for Set2p in transcriptional elongation. *SET2* deletion has been shown to be synthetically sick with deletions of other elongation factors (Krogan et al., 2003).

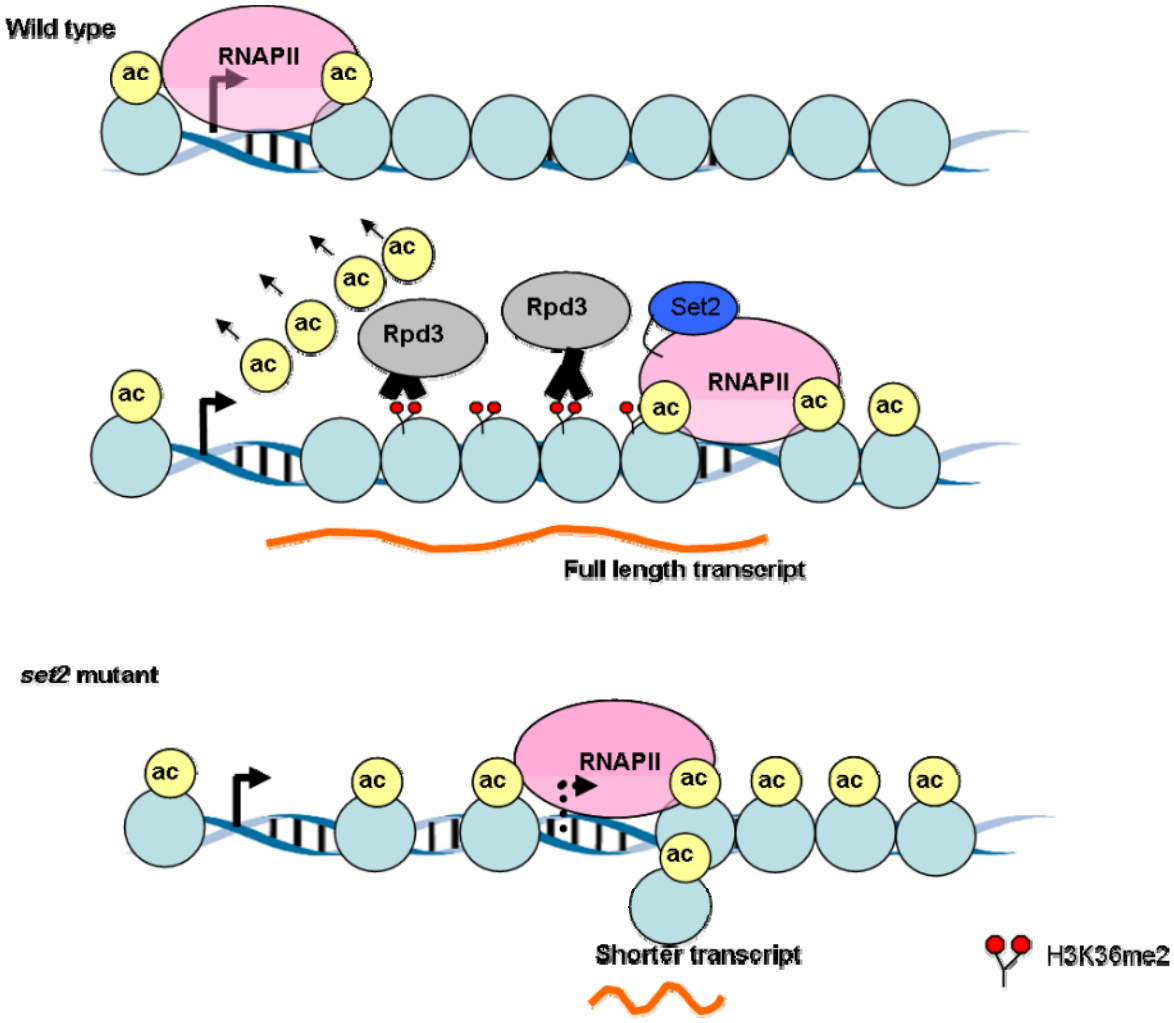


(Figure courtesy of Dr. Brian Strahl)

Figure 1 Set2p associates with the Ctk1p-phosphorylated C-terminal domain of RNA pol-II and this association is required for methylation at histone H3K36

H3K36me has been shown to recruit histone deacetylase complex Rpd3(S) (Carrozza et al., 2005; Joshi and Struhl, 2005; Keogh et al., 2005). The Rpd3(S) complex binds to H3K36me via Eaf3, a methyllysine binding chromodomain protein. Deacetylation of histones is required to maintain nucleosome integrity in the wake of the passing polymerase and this is directed by H3K36me. These studies show increased acetylation in the 3' region of genes in the *set2* and *eaf3* deletion strains. (Carrozza et al., 2005) showed that in the absence of *SET2* and subunits of Rpd3(S) complex genes exhibited aberrant transcription initiation from within the coding region of the genes thus suggesting that H3K36me is important for ensuring that initiation occurs at the

5’ end of genes by maintaining proper acetylation levels in the coding regions. H3K36me, therefore, is a mechanism with which chromatin regulates transcription fidelity.



(Model based on the results from Carrozza et al, 2005; Keogh et al, 2005; Joshi and Struhl, 2005)

Figure 2 Deletion of Set2p results in aberrant transcription initiation.

Studies showed that the histone deacetylase complex, Rpd3(S), binds to H3K36me and deacetylates nucleosomes in the wake of the passing polymerase during transcription elongation. Thus, Set2p-mediated Rpd3(S) recruitment maintains proper acetylation levels within the ORFs, the loss of which results in hyperacetylation and aberrant initiation

Chapter II

DIMETHYLATION OF HISTONE H3 AT LYSINE 36 DEMARCATES REGULATORY AND NON-REGULATORY CHROMATIN GENOME-WIDE

This work was published as:

Bhargavi Rao, Yoichiro Shibata, Brian D. Strahl and Jason D. Lieb

Molecular and Cellular Biology, 2005 Nov; 25(21):9447-59

ABSTRACT

Set2p, which mediates histone H3K36me2 in *Saccharomyces cerevisiae*, has been shown to associate with RNA pol-II at individual loci. Here, ChIP-chip experiments normalized to general nucleosome occupancy reveal that nucleosomes within ORFs and downstream non-coding chromatin were highly dimethylated at H3K36, and that Set2p activity begins at a stereotypic distance from the initiation of transcription genome-wide. H3K36me2 is scarce in regions upstream of divergently transcribed genes, telomeres, silenced mating loci, and regions transcribed by RNA pol-III, providing evidence that the enzymatic activity of Set2p is restricted to its association with RNA pol-II. The presence of H3K36me2 within ORFs correlated with the "on" or "off" state of transcription, but the degree of H3K36 dimethylation within ORFs did not correlate with transcription frequency. This provides evidence that H3K36me2 is established during the initial instances of gene transcription, with subsequent transcription having at most a maintenance role. Accordingly, newly activated genes acquire H3K36me2 in a manner that does not correlate with gene transcript levels. Finally, nucleosomes dimethylated at H3K36 appear to

be refractory to loss from highly transcribed chromatin. Thus H3K36me2, which is highly conserved throughout eukaryotic evolution, provides a stable molecular mechanism for establishing chromatin context throughout the genome by distinguishing potential regulatory regions from transcribed chromatin.

INTRODUCTION

In eukaryotic cells, the accessibility of the DNA template is influenced by chromatin structure. For example, in *Saccharomyces cerevisiae*, transcription factors have been shown to bind to consensus sequences upstream of genes in preference to identical consensus sequences that occur within the coding sequences of transcribed genes (Kuo et al., 2000; Lieb et al., 2001). Likewise, transposons preferentially insert into promoter regions (Mai et al., 2000), and the double-strand breaks required for meiotic recombination in *S. cerevisiae* occur preferentially in gene promoters rather than in the coding regions (Gerton et al., 2000; Wu and Lichten, 1994). Chromatin context therefore is a major determinant of where on the genomic DNA template many biological phenomena will occur.

Regulation of accessibility to the DNA template is likely to be mediated in large part through differential regulation of nucleosome occupancy. Promoter regions of *S. cerevisiae* exhibit reduced nucleosome occupancy genome-wide (Bernstein et al., 2004; Lee et al., 2004), and these differences in nucleosome occupancy are important for promoter accessibility (Sekinger et al., 2005a). Furthermore, in *S. cerevisiae*, promoter and nonregulatory chromatin can be biochemically fractionated, indicating that those regions have distinct physical properties (Nagy et al., 2003). Nucleosomes can be moved or displaced from specific genomic regions by several general mechanisms, including nucleosome-remodeling complexes like SWI/SNF and RSC

(Martens and Winston, 2003), binding of activators to DNA (Boeger et al., 2003; Boeger et al., 2004; Morse, 2000), transcriptional elongation by RNA pol-II (Kristjuhan and Svejstrup, 2004; Lee et al., 2004; Schwabish and Struhl, 2004), and inherent properties of DNA sequence (Sekinger et al., 2005b). Template accessibility and nucleosome occupancy can also be mediated by posttranslational modification of the N-terminal histone tails, most notably acetylation (Lee et al., 1993; Sewack et al., 2001). Although chromatin context may be defined in part by regional differences in histone modifications, no chromatin mark has been shown to correspond specifically to coding or regulatory regions throughout the genome. Here, we present evidence that H3K36me2, which is mediated by the methyltransferase Set2p (Landry et al., 2003; Strahl et al., 2002), may provide such a mark.

Set2p interacts with the CTD of RNA pol-II (Krogan et al., 2003; Li et al., 2002; Xiao et al., 2003), and this interaction is regulated by the phosphorylation state of the CTD. Ser5 of the CTD repeat is phosphorylated by Kin28p during initiation of transcription, while Ser2 and Ser5 are phosphorylated by Ctk1p during elongation (Cho et al., 2001; Jones et al., 2004; Komarnitsky et al., 2000; Licatalosi et al., 2002). Set2p associates preferentially with Ser2/Ser5 phosphorylated repeats of the RNA pol-II CTD, and deletion of *CTK1* abolishes H3K36me2 (Krogan et al., 2003; Xiao et al., 2003). Set2p-RNA pol-II interactions are also dependent on the Paf1 complex (Paf1p, Rtf1p, Cdc73p, Ctr9p, and Leo1p) (Krogan et al., 2003), which also associates with RNA pol-II (Krogan et al., 2002; Pokholok et al., 2002; Squazzo et al., 2002). This and other biochemical data suggest that Set2p associates with RNA pol-II specifically during transcription elongation (Kizer et al., 2005; Krogan et al., 2003; Li et al., 2002; Schaft et al., 2003; Xiao et al., 2003). Chromatin immunoprecipitation (ChIP) assays followed by quantitative PCR on a few selected loci have supported this assertion, showing that H3K36me2 is generally restricted to the

transcribed regions of RNA pol-II regulated genes (Bannister et al., 2005; Kizer et al., 2005; Krogan et al., 2003; Schaft et al., 2003).

While there is strong evidence that Set2p is associated with elongating polymerase, the physiological functions of Set2p and H3K36me2 are still unknown. Evidence suggesting a function for Set2p in transcriptional elongation comes from results showing either sensitivity or resistance of *set2*Δ strains to the elongation inhibitor 6-azauracil. These phenotypes are similar to those exhibited by strains defective for genes encoding elongation factors like Chd1p, Iswi1p, and Fkh1p (Krogan et al., 2003; Li et al., 2003; Li et al., 2002; Schaft et al., 2003; Xiao et al., 2005). A role in transcriptional elongation is also supported by synthetic genetic interactions between *set2*Δ and deletions of all members of the Paf1 complex, the chromodomain factor Chd1p, a putative elongation factor Soh1p, and the Bre1p or Lge1p components of histone H2B ubiquitination complex (Krogan et al., 2003). However, whatever role Set2p plays in elongation is either not essential or redundant, since *set2*Δ strains are viable and, in many backgrounds, exhibit very mild phenotypes.

To further elucidate the cellular function of H3K36me2, we determined its pattern of distribution throughout the *S. cerevisiae* genome. We performed additional experiments to determine how the pattern of H3K36me2 changes in response to a change in global transcriptional state and the relationship between the H3K36me2 mark and nucleosome stability. H3K36me2 demarcates the structurally distinct regulatory and nonregulatory regions of yeast genomic chromatin and may serve as an indicator of chromatin context.

MATERIALS AND METHODS

Strains and Culture Conditions. For H3K36me2 and histone H3 ChIPs, strain AS4 (*MAT alpha, trp1-1, arg4-17, tyr7-1, ade6, ura3*) was used (Stapleton and Petes, 1991). For histone H4 ChIPs, a previously described myc-tagged H4 strain constructed in UCC1111 [(*MAT alpha, ade2::his3-Δ200, leu2-Δ0, lys2-Δ0, met15-Δ0, trp1-Δ63, ura3-Δ0, adh4_URA3-TEL (VII-L), hhf2-hht2::MET15, hhf1-hht1::LEU2, pRS412 (ADE2 CEN ARS)-HHF2-HHT2]*) was used (Ng et al., 2003a; Ng et al., 2002). Unless otherwise described, yeast was grown to an OD_{600} of 0.8-1.0 with shaking at 30^{o}C in 100ml of YPD media (1% yeast extract, 2% peptone, 2% dextrose).

Antibodies. Antibodies against histone H3 lysine 36 dimethylation have been previously described (Strahl et al., 2002) and were derived from Upstate (Cat # 07-369). Myc-tag antibodies were also obtained from upstate (Cat # 05-419). The histone H3 is a rabbit polyclonal to human histone H3 amino acid corresponding to 124-135 (CGIQLARRIRGERA) obtained from Abcam Inc. (AB1791).

Dot Blot. Peptides (KSAPSTGGV<u>K</u>KPHRYKPGTGK-BIOTIN) in which the residue corresponding to H3K36 (underlined) was either mono-, di-, or trimethylated were resuspended in ddH_2O (10 μg/μL) and serially diluted in TBS (150 mM NaCl, 10 mM Tris pH 7.6). Aliquots of 100-μL peptide/TBS solution were spotted onto PVDF using a Bio Rad dot blot apparatus. Membranes were washed in TBS and then blocked in 2.5% TBS-T (TBS, 0.1% Tween 20), for 10 minutes prior to incubation with a 1:10,000 dilution of the specified antibody for 2 hrs at room temperature. Membranes were washed with TBS-T for 10 minutes three times, incubated

with anti-rabbit HRP-conjugated IgG for two hours at room temperature, and then washed again for 10 minutes three times prior to detection using ECL-Plus from Amersham.

Chromatin Immunoprecipitation (ChIP) assays. ChIP assays were performed as described (Kuo and Allis, 1999). Briefly, whole cell extracts were prepared from 1% formaldehyde-fixed wildtype and *set2Δ* cells using lysis buffer (50 mM Hepes-KOH, pH 7.5, 300 mM NaCl, 1 mM EDTA, 1% Triton-X, and 0.1% sodium deoxycholate) and sonicated to shear the chromatin (0.25-1-kb range). Immunoprecipitation was performed with anti-H3K36me2, anti-myc or anti-H3. After cross-link reversal at 65^0C, DNA was extracted according to the manufacturer's instructions using the Qiagen PCR-purification kit.

DNA Amplification, Labeling, Array Hybridization and Data Processing. ChIP-enriched DNA and reference DNA in all experiments was amplified as described (Bohlander et al., 1992). Briefly, two initial rounds of DNA synthesis with T7 DNA polymerase using primer 1 (5′-GTTTCCCAGTCACGATCNNNNNNNNN-3′) was followed by 25 cycles of PCR with primer 2 (5′-GTTTCCCAGTCACGATC-3′). Cy3-dUTP or Cy5-dUTP were then incorporated directly with an additional 25 cycles of PCR using primer 2. Microarray hybridizations were performed using standard procedures (Iyer et al., 2001). The arrays were scanned with a GenePix 4000 scanner and data were extracted with Genepix 5.0 software. Data were normalized such that the median $\log_2$ ratio value for all quality elements on each array equaled zero, and the median of pixel ratio values was retrieved for each spot. Only spots of high quality by visual inspection, with at least 50 pixels of quality data (regression $R^2 > 0.6$), and for which intensity of the reference signal was strong (>350 units) were used for analysis. Arrayed elements that did not

meet these criteria on at least half of the arrays were excluded from analysis. All data were log-transformed before further analysis. For normalization with the nucleosome occupancy data, the median $\log_2$ ratio values of H4-myc ChIP were subtracted from the median H3K36me2-ChIP ratio values. Unless otherwise noted, all data presented is nucleosome-occupancy normalized in this way. While many methods of bulk-nucleosome normalization are possible, all must contend with the inherent difficulties of combining ChIP datasets produced with two different antibodies (Buck and Lieb, 2004). The method used here is simplest, and provides a more realistic representation of the modification pattern than does unnormalized data. I provide all raw data (see below) so that readers may apply their preferred normalization method.

DNA Microarray Preparation. ORFs and intergenic regions from yeast (S288C) were PCR-amplified and printed on poly-lysine coated glass slides using a robotic arrayer as described (Iyer et al., 2001). ORFs were generally represented by PCR products that extended from start codon to stop codon. Elements representing intergenic regions generally included all DNA between annotated ORFs, with the fragments divided such that PCR products were no longer than 1.5 kb.

Locus-specific detection of ChIP enrichment. The sequences of the primers used in Figure 4 and Figure 8 are shown in Table 1.

Data Availability. All raw microarray data and images are available to the public through the UNC microarray database (https://genome.unc.edu/). Data is also available from a web supplement at (https://genome.unc.edu/pubsup/H3k36me2/) and through GEO (accession numbers GPL2503-GPL2506).

RESULTS

Chromatin from ORFs and regions immediately downstream of ORFs is enriched for histone H3 lysine 36 dimethylation. As the initial step in my goal to determine the genome-wide location of H3K36me2 in *S. cerevisiae*, we characterized a polyclonal antibody directed against H3K36me2. While the general specificity of this antibody for methylation at H3K36 had been previously verified (Kizer et al., 2005; Xiao et al., 2005), its precise specificities to the different possible H3K36 methylation states (mono-, di-, and trimethylation) were unknown. To determine the specificity of this antibody for H3K36 methylation, dot blots were performed by my colleague, Dr. Ronald Laribee in Dr. Brian Strahl's lab against peptides that were either mono-, di-, or tri-methylated at the residue corresponding to H3K36 (see Materials and Methods). As shown in Fig. 3A, the antiserum was specific to dimethylation of H3K36 and did not cross-react with any of the related modifications.

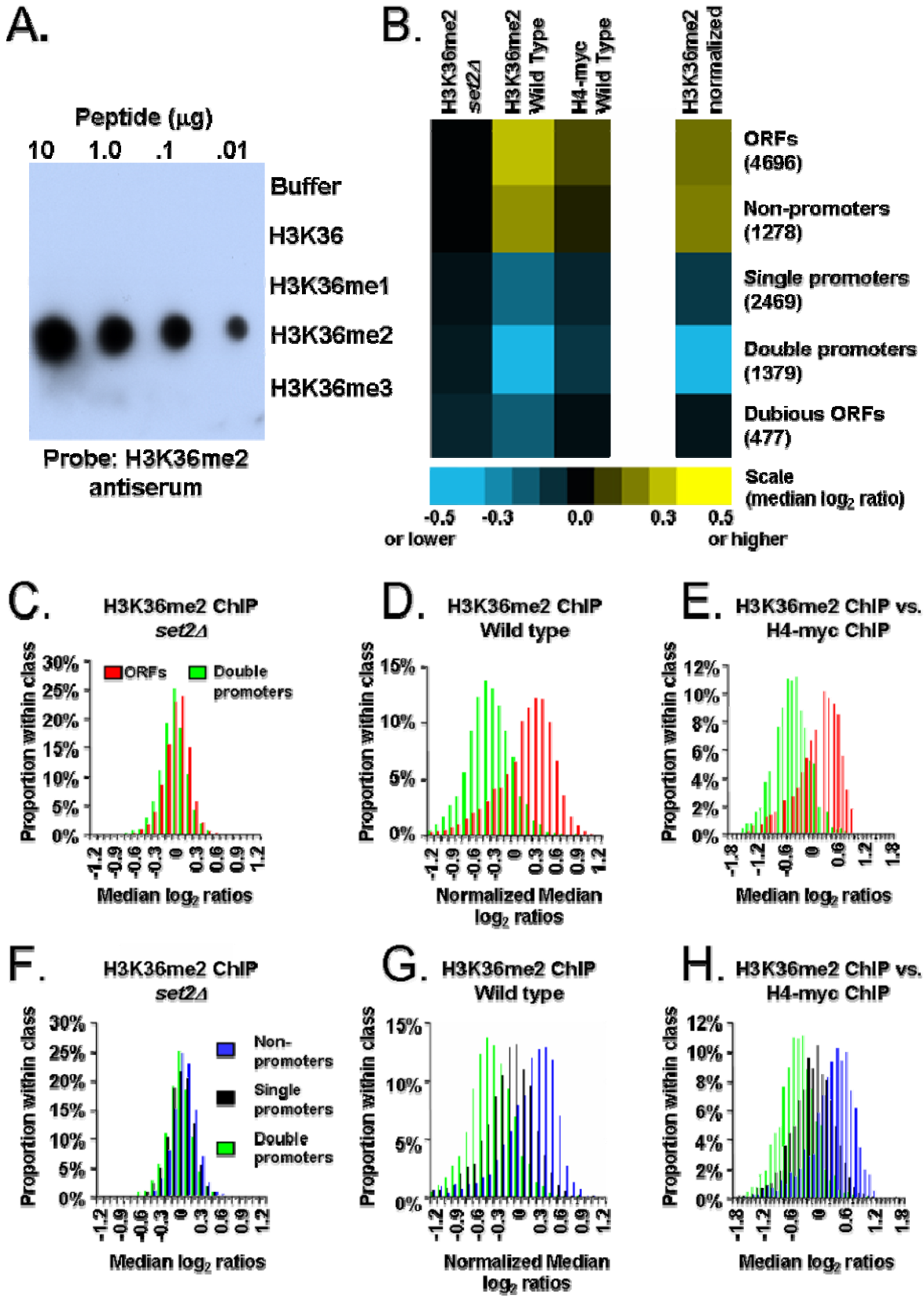

A.
Peptide (μg)
10 1.0 .1 .01
Buffer
H3K36
H3K36me1
H3K36me2
H3K36me3
Probe: H3K36me2 antiserum
B.
H3K36me2 set2Δ
H3K36me2 Wild Type
H4-myc Wild Type
H3K36me2 normalized
ORFs (4696)
Non-promoters (1278)
Single promoters (2469)
Double promoters (1379)
Dubious ORFs (477)
Scale (median log2 ratio)
-0.5 -0.3 0.0 0.3 0.5
or lower
or higher
C.
H3K36me2 ChIP
set2Δ
ORFs
Double promoters
Proportion within class
Median log2 ratios
D.
H3K36me2 ChIP
Wild type
Normalized Median log2 ratios
E.
H3K36me2 ChIP vs. H4-myc ChIP
F.
Non-promoters
Single promoters
G.
H.

Figure 3 H3K36me2 is restricted to RNA pol-II transcribed regions genome-wide.

(A) The indicated amounts (top) of the specified peptides (right) were blotted onto a polyvinylidene difluoride membrane and probed with a 1:10,000 dilution of the H3K36me2 antiserum. To verify the presence of the H3K36me1 and H3K36me3 peptides, antibodies specific to each of those peptides were used in parallel to probe identical blots. Experiments performed by Nick Laribee. (B) Colors (scale at bottom) represent the median of ratios [log2 (H3K36me2 ChIP signal intensity/normalized reference signal intensity)] recorded from all arrayed elements in the indicated Saccharomyces Genome Database functional class (numbers of arrayed elements in parentheses). Data were derived from 12 independent wild-type and 8 set2Δ H3K36me2 ChIP experiments (biological replicates). Intergenic regions are organized into the following three categories: double, upstream of two divergent genes; single, upstream of one gene; non, upstream of zero genes. (C) A histogram showing the distribution of H3K36me2 ChIP median log2 ratios in a set2Δ strain. Ratios in panels C to H were normalized to bulk nucleosome occupancy by subtracting median H4-myc log2 ratios (eight independent experiments, wild type) or anti-H3 ChIPs (five independent experiments, set2Δ) from median H3K36me2 ChIP log2 ratios (same 12 ChIPs as mentioned for panel B). (D) Same as panel C, but for a wild-type strain. (E) Same as panel D, but instead of computational normalization, plotted ratios were derived from direct hybridization of H3K36me2 ChIP versus H4-myc ChIP (four independent ChIP sets). Note that for panels C to E, a positive correlation between the ratios reported at double promoters and the ratios reported at adjacent ORFs was observed (data not shown). Therefore, any high ratios reported at double promoters may, at least in part, be attributed to high ratios in adjacent ORFs. This is likely due to the limited resolution of the ChIP-chip procedure and my microarrays. (F) Same as panel C, but for noncoding regions. (G) Same as panel D, but for noncoding regions. (H) Same as panel E, but for noncoding regions.

Having verified the specificity of this antiserum, ChIP experiments were performed using extract from wild-type strains. To assess the relative abundance of genomic fragments enriched by the ChIP, samples were RNase treated and DNA was amplified and labeled fluorescently. In parallel, total genomic DNA was prepared from input extract, RNase treated, amplified, and labeled with a different fluorescent marker. The two samples were then analyzed by comparative hybridization to DNA microarrays. The microarrays used in this study cover the entire yeast genome, including the coding and noncoding regions, at approximately 1-kb resolution (see Materials and Methods). H3K36me2 ChIP-chip experiments were performed with a total of 12 independent wild-type yeast cultures. As a control, ChIP-chip experiments were performed from each of eight independent extracts in which H3K36me2 was eliminated by deletion of the *SET2* gene. I found that H3K36me2 ChIPs enriched chromatin corresponding to ORFs relative to chromatin from genomic regions upstream of genes (.Figure 3B).

In *S. cerevisiae*, the noncoding regions downstream of two convergently transcribed genes are almost always completely transcribed, often on both strands, by the converging polymerases (Hurowitz and Brown, 2003). I found that these regions, which correspond to 3' untranslated regions (UTRs), were enriched by H3K36me2 ChIPs at a level equal to or greater than the enrichment observed at ORFs (Figure 3B). To confirm that my ChIPs were reflections of H3K36me2 levels, I performed ChIP experiments using extracts from *set2*Δ strains. Very little DNA was recovered from these ChIPs, and analysis of the DNA that was recovered revealed none of the specific patterns described above (Figure 3B). I therefore interpret the efficiency of DNA recovery at each locus after H3K36me2 ChIP to reflect relative H3K36me2 levels. The evidence presented thus far supports the hypothesis that regions of the genome transcribed by RNA pol-II are enriched for H3K36me2.

Chromatin upstream of ORFs is H3K36me2 deficient. In further support of the hypothesis that H3K36me2 is restricted to transcribed regions, the lowest levels of H3K36me2 were found in chromatin upstream of two divergently transcribed genes ("double promoters"), which is not expected to be transcribed by RNA pol-II (Figure 3B). On the other hand, "single promoters" are expected to be partially transcribed since they contain the 3' UTR of the upstream gene. As predicted, single promoters exhibit a level of enrichment lower than that observed for ORFs and 3' UTRs but higher than that observed for double promoters (Figure 3B). These experiments provide evidence that dimethylation of histone H3 at lysine 36 is absent from regions of the genome that are not transcribed by RNA pol-II.

The genomic pattern of H3 lysine 36 dimethylation persists after normalization for general nucleosome occupancy. Nucleosome occupancy is generally lower in noncoding regions

upstream of genes than in ORFs (Bernstein et al., 2004; Boeger et al., 2003; Boeger et al., 2004; Lee et al., 2004; Reinke and Horz, 2003; Svaren and Horz, 1997). Thus, we wondered if the pattern I observed with the H3K36me2-specific antiserum was a reflection, at least in part, of general nucleosome occupancy. To ensure that my results were specific to the H3K36me2 modification, I normalized my H3K36me2 distribution data to general nucleosome occupancy. I prepared extracts from yeast cells in which the only source of histone H4 was tagged with the myc epitope and performed ChIP assays using anti-myc antibodies. Histone H4 ChIPs were performed on five independent yeast cultures. Consistent with published data (Lee et al., 2004), results of the histone occupancy ChIPs revealed that nucleosomes were more enriched in the coding region of genes than in intergenic regions (Figure 3B). Indistinguishable results were obtained with nucleosome ChIP-chips using an antibody specific to the C terminus of histone H3 (data not shown). In parallel, H3K36me2 ChIPs were performed using the same extracts. Even without normalization, the qualitative differences between the distribution of H3K36me2 and general nucleosome occupancy indicated that the H3K36me2 pattern was indeed distinct. Specifically, noncoding regions downstream of convergently transcribed genes were enriched by the H3K36me2 ChIP at a level nearly equal to the enrichment observed at ORFs, whereas in histone H3 or H4 ChIPs, ORFs were more strongly enriched than 3' UTRs (Figure 3B).

For further data analysis, we chose the simplest possible normalization routine by subtracting the median $\log_2$ ratio values of the H4-myc ChIP-chip data from the median ratio values of H3K36me2 ChIP-chip data (see Materials and Methods). After normalization, the clear enrichment of transcribed genomic regions and corresponding depletion of regulatory regions of the genome persisted (Figure 3C, D, F and G). As a test of the validity of this normalization

approach, I performed direct comparative hybridizations between DNA enriched by H3K36me2 ChIP and DNA enriched with H4-myc ChIP. The data obtained from direct comparative hybridizations were essentially identical to the computationally normalized H3K36me2 data (Figure 3E and H).

Higher resolution, locus-specific ChIPs confirm that H3K36me2 is concentrated in chromatin within coding regions and at the 3' end of genes. In H3K36me2 ChIPs, chromatin downstream of convergently transcribed genes was by far the most highly enriched class of noncoding chromatin (Figure 3G). This suggested that Set2p is active throughout the entire transcript length, providing a possible mechanism for distinguishing nonregulatory intergenic regions from promoters. To validate this observation, Yoichiro Shibata, my collaborator in Dr. Brian Strahl's lab interrogated the ChIP results with PCR primers that represent regulatory and transcribed chromatin across a 16-kb region on chromosome XII (Figure 4). He found that the chromatin heavily enriched by H3K36me2 ChIPs corresponded to coding regions and to regions lying downstream of two convergently transcribed genes. For example, *SST2* and *LEU3* are both transcribed under the conditions assayed. Their 3' UTRs are each about 450 bp in length (Hurowitz and Brown, 2003) and are represented by the primer sets B, C, and D in Figure 4. Chromatin covered by these primer sets is among the most heavily enriched in the tested region.

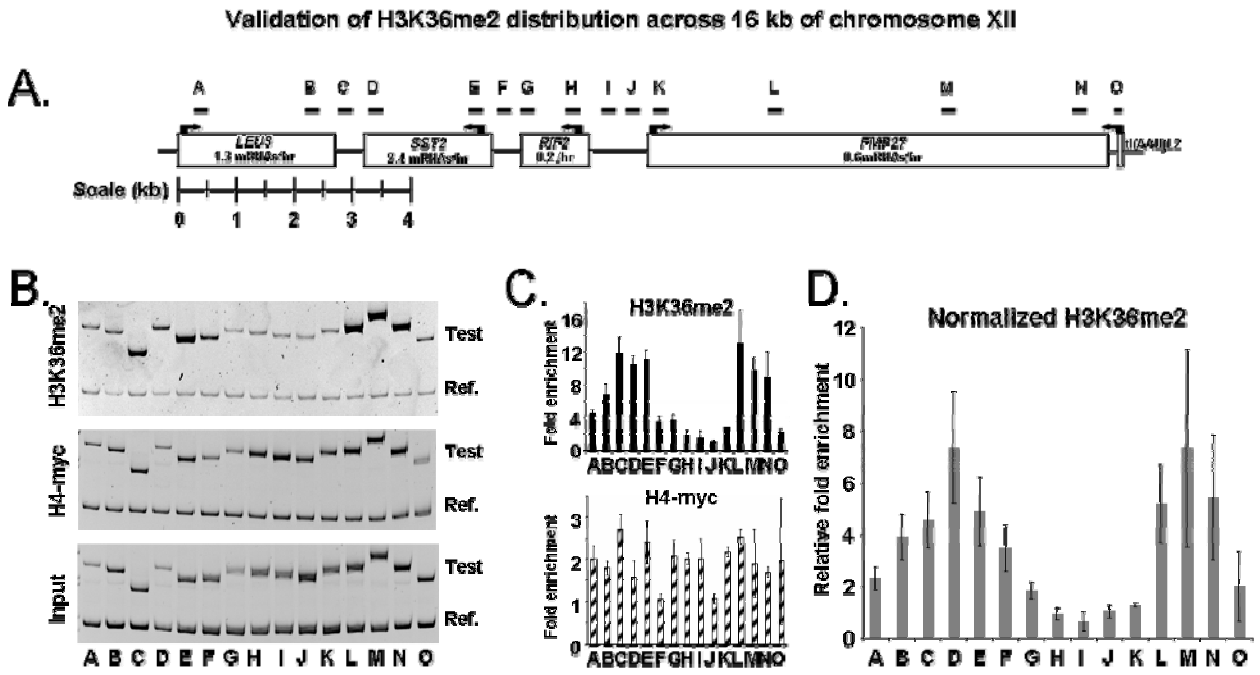


Figure 4 Validation and fine-scale mapping of H3K36me2 distribution across 16 kb of chromosome XII.

(A) Schematic of chromosome XII coordinates 1,036,089 to 1,052,141, showing the locations of primers used to interrogate three of the independent H3K36me2 ChIP-chips shown in Fig. 1.1. Primer sets are listed in Table 1. (B) Polyacrylamide gel analysis of PCR products generated by the primer sets shown in panel A (Test). The reference product (Ref.) corresponds to 146 bp of a large noncoding region between YEL073C and YEL072W on chromosome V. (C) Quantitation of the gel shown in panel B. Graphs show the average enrichments detected by the PCR primer sets following H3K36me2 and histone H4-myc ChIPs. The value plotted for each fragment was calculated as follows: [(w/x)/(y/z)], where all values are the sum of pixel intensities for each band. w, ChIP fragment; x, ChIP reference; y, input test fragment; z, input reference. Numbers therefore reflect relative immunopreciptation efficiencies; values of less than 1 may be expected. Error bars illustrate the average deviations from the means. (D) H3K36me2 enrichment values after normalization with general nucleosome occupancy levels. The values plotted were calculated as follows: [(w/x)/ (y/z)]H3K36me2/[(w/x)/(y/z)]H4-myc. These experiments were performed by Yoichiro Shibata. Figure courtesy of Yoichiro Shibata.

Levels of histone H3K36me2 do not correlate with transcription frequency. Localization of H3K36me2 to chromatin in the body of the RNA pol-II transcribed genes is consistent with the earlier studies showing that Set2p associates with the elongating form of RNA pol-II. I wondered if the frequency of transcription correlated with the degree of H3K36me2. The transcription frequency (also called transcription rate) for each *S. cerevisiae* gene has been calculated based on

measurements of steady-state RNA levels and RNA half-lives in exponentially growing yeast cells at 30°C (Holstege et al., 1998). I compared these published transcription frequency values to the results of 12 independent H3K36me2 ChIP-chip experiments. I found that among genes with measurable transcription frequencies (>0 mRNA/hour), the level of H3K36me2 enrichment did not correlate with transcription frequency (Figure 5A). Genes with transcription frequencies ranging from 1 to 120 mRNAs/hour were consistently enriched in the H3K36me2 ChIPs. For example, despite low rates of transcription, genes like *BUD14* and *TPK2* (both 1.8 mRNAs/hour) were enriched in H3K36me2 ChIPs (97th and 95th ChIP percentiles, respectively) as highly as were heavily transcribed genes like *HXK2* (71 mRNAs/hour, 96th ChIP percentile). These results suggest that H3K36me2 occurs chiefly in the initial instance or early instances of gene transcription, with subsequent transcription playing at most a maintenance role.

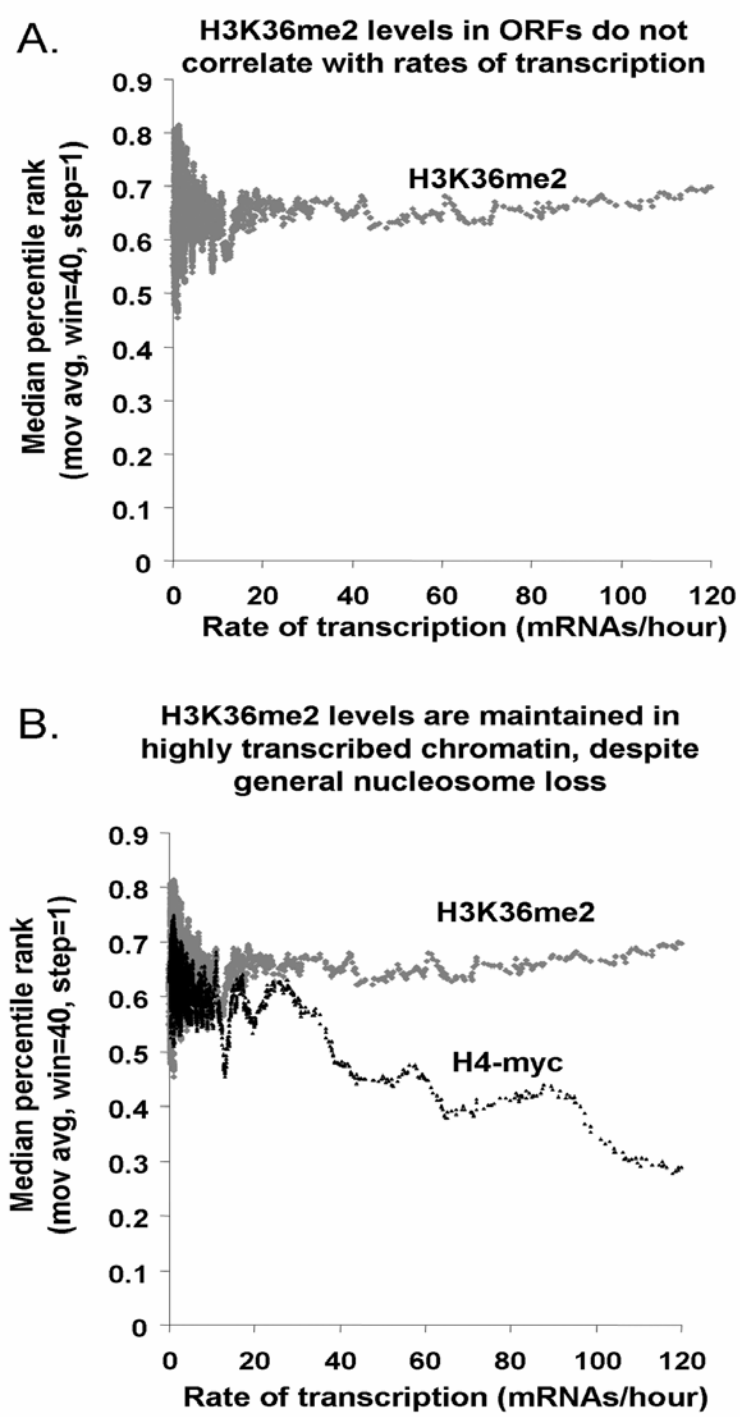


Figure 5 H3K36me2 levels in ORFs do not correlate with transcription frequencies.

(A) Genes were sorted by their transcription rates during mitotic growth (x axis) (Holstege et al., 1998), and a moving average (mov avg) (window [win] = 40, step = 1) of the degrees of their enrichment in H3K36me2 ChIPs was plotted (percentile rank, y axis). In this plot, H3K36me2 ChIP values are not

normalized to overall nucleosome occupancy. (B) Same as panel A, except that both H3K36me2 ChIPs and H4-myc ChIPs are plotted.

H3K36me2 levels appear to compensate for transcription-coupled nucleosome loss. Previous studies have shown that chromatin in very highly transcribed genes exhibits relatively low nucleosome occupancy, suggesting that nucleosomes are either removed or displaced temporarily on transcriptionally active chromatin (Lee et al., 2004; Schwabish and Struhl, 2004). However, we did not observe this phenomenon for nucleosomes that contain dimethylated H3K36. The levels of H3K36me2 appeared to remain constant on genes, irrespective of their rates of transcription (Figure 5B). This suggests that the level of H3K36me2 is maintained on highly transcribed chromatin either by preferential retention of nucleosomes containing H3K36me2 or by nonlinear increases in H3K36 methylation as a function of transcription rate. For the "nonlinear" hypothesis to be true, the H3K36me2 mark would have to be less than saturated at all genes, and increased enzymatic activity of Set2p would have to be linked to increased transcriptional activity for only the subset of very heavily transcribed genes at which bulk nucleosome loss has been observed. Since no correlation between transcription rate and H3K36me2 level is observed across a broad range of transcription rates (Figure 5A), we favor the "preferential retention" hypothesis.

We then asked whether nucleosomes are less stable on highly transcribed chromatin in the absence of H3K36me2. I measured bulk nucleosome occupancy by performing histone H3 ChIPs in a *set2*Δ strain. However, I found that nucleosomes are lost from highly transcribed chromatin equally in *set2*Δ and wild-type strains, indicating that nucleosome occupancy is not directly affected by H3K36me2 (Figure 6). Therefore, additional mechanisms may work to stabilize

H3K36me2 nucleosomes, or H3K36me2 may be an indicator, but not a cause, of transcription-stable nucleosomes.

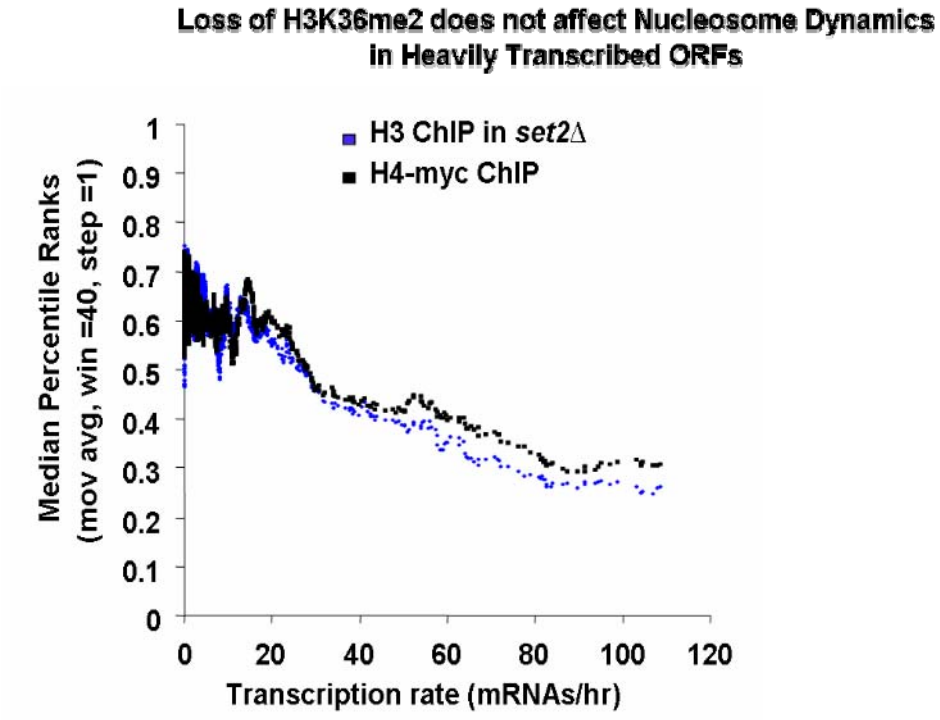


Figure 6 The extent of loss of nucleosomes in *set2*Δ is same as in wildtype.

Moving average of median percentile ranks (window [win] = 40, step = 1) of ChIP enrichment values for ORFs, plotted as a function of transcription rate (mRNAs/hr). H4-myc ChIPs (black) in a wild-type strain; H3 ChIPs in a *set2*Δ strain (blue).

H3K36me2 correlates with "on" or "off" transcriptional state. The hypothesis that H3K36me2 is stable and occurs chiefly in the initial instance of gene transcription predicts no correlation between H3K36me2 level and transcriptional rate (as I observed) but does predict a positive correlation between H3K36me2 level and the on or off transcription state of a gene. We defined a gene as on if it had a measurable transcription rate (>0 mRNA/hour) and off if it did not (0 mRNA/hour) (Holstege et al., 1998). To test this prediction, I ranked the ORFs according

to their enrichment levels in H3K36me2 ChIP experiments and divided them equally into 10 bins, such that the least enriched 10% of ORFs were in bin 1, the most enriched 10% were in bin 10, and so on. I then simply asked what proportion of genes in each bin was on (Figure 7A). The results show that genes that were not enriched by H3K36me2 ChIPs were more likely than others to be off and that the likelihood of any given gene to be off decreased with increasing H3K36me2 ChIP enrichment. No such trend was observed with H4-myc ChIPs (Figure 7A) or with H3K36me2 ChIPs performed with a *set2*Δ strain (data not shown). This result provides evidence that H3K36me2 is not linked with how often a gene is transcribed per se but rather with the occurrence of transcription.

Dormant chromatin that becomes transcriptionally active acquires H3K36me2. The hypothesis that H3K36 dimethylation is established by initial instances of gene transcription predicts that upon a switch from an inactive to an active state, chromatin will become dimethylated at H3K36. To test this prediction, I induced transcription at hundreds of genes simultaneously by subjecting yeast cells to transfer from 25°C to 37°C (Causton et al., 2001; Gasch et al., 2000). The level of Set2 protein levels did not change after the heat shock (data not shown).

ORFs that become transcribed during heat shock acquire the H3K36me2 mark, while a relative decrease in H3K36me2 levels is observed in ORFs that are repressed (Figure 7B). Of particular interest are the genes that were off during log phase but strongly induced after heat shock (measured expression increase of $>\log_2 2$). Of these genes, 75% (48/65) had increased levels of H3K36me2. In contrast, among genes that were off during log phase and remained off during

heat shock (expression increase of $<\log_2 1$), only 47% (144 of 306) exhibited increased levels of H3K36me2. The difference between these groups was significant (χ^2 test; $P = 8.6 \times 10^{-5}$).

I also observed a relative loss of H3K36me2 in the ORFs of repressed genes. Of genes that were on during log phase and that remained active ($\log_2$ expression ratios of >-1), only 46% (1,099/2,388) had increased H3K36me2 levels. In contrast, of genes that were on during log phase and repressed fourfold or more after heat shock, 65% (210/332) had decreased levels of H3K36me2 (χ^2 test; $P = 0.0014$). By examining H3K36me2 ChIP data that were not normalized to nucleosome occupancy (data not shown), we found that this relative decrease is attributable in part to bulk nucleosome replenishment at repressed genes (Lee et al., 2004) rather than loss of H3K36me2 on existing nucleosomes.

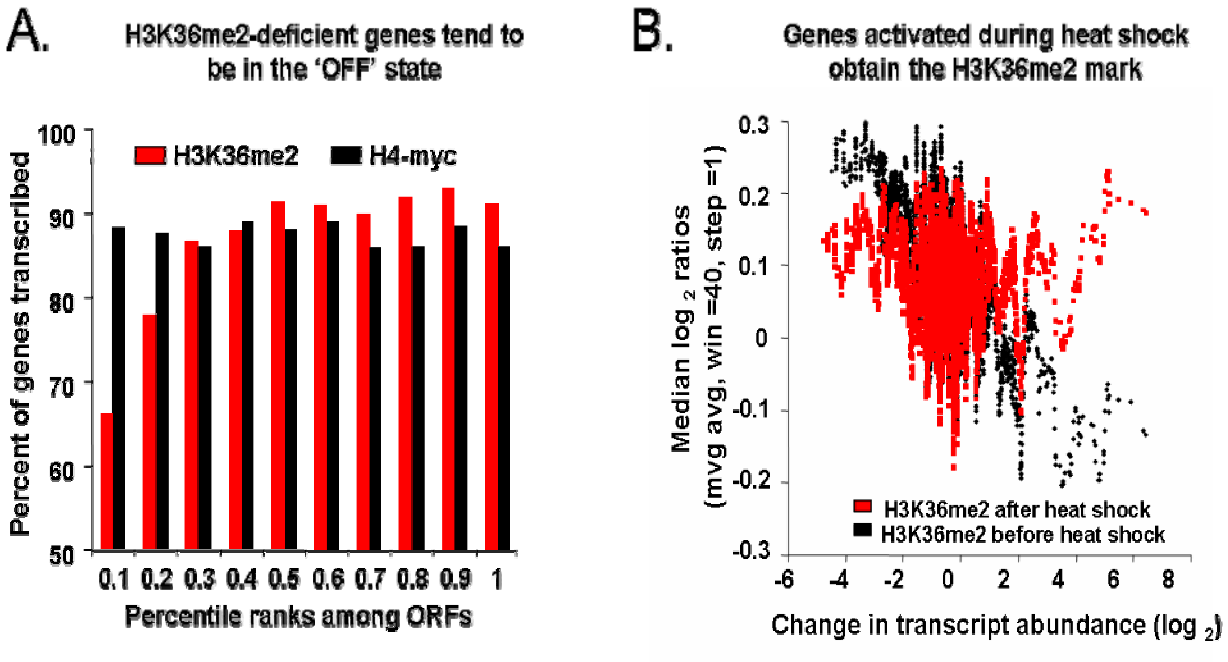


Figure 7 H3K36me2 levels correlate with "on" or "off" transcriptional state.

(A) ORFs were divided equally into 10 bins according to their degrees of enrichment in H3K36me2 ChIPs (normalized to H4-myc) or H4-myc ChIPs. The ORFs in each bin were then classified as either "on" (>0 mRNAs/hr) or "off" (0 mRNAs/hr). The percentage of genes in each bin that was classified as "on" is shown. We note that ORF size and transcription state are not independent variables, since shorter ORFs tend to be in the off state, even after removal of SGD-annotated dubious ORFs. However, analysis of ORFs of similar size rather than all ORFs reveals the same pattern shown here. Note that ~85% of genes are transcribed at detectable levels during mitotic growth (Holstege et al., 1998; Hurowitz and Brown, 2003), so the uniform H4-myc data reflect a neutral relationship with transcription state. (B) Genes were sorted according to the degrees of their transcript level change upon heat shock (*x* axis) (Gasch et al., 2000). A moving average (mov avg) (window [win] = 40, step = 1) of relative H3K36me2 ChIP enrichment (normalized to general histone H4 occupancy measured in parallel) at each ORF (*y* axis) is plotted before and after heat shock. Most genes are unchanged upon heat shock and are therefore clustered near the center of the graph. Points to the right represent ORFs that are activated by heat shock, and those to the left represent ORFs repressed by heat shock.

To further confirm my ChIP-chip data, my collaborator at the Strahl lab, Yoichiro Shibata, performed quantitative ChIP analysis before and after heat shock along the length of *PHM7* (Fig. 2.6A). *PHM7* is repressed in logarithmically growing cell cultures (~0 mRNA per hour) but is induced by a factor of 7 during heat shock (Gasch et al., 2000). The results showed that this gene

acquired the H3K36me2 mark after heat shock and only in the 3' region of the ORF (primer sets R and S) (Figure 8B to C). This result persisted after normalization to bulk histone occupancy changes following heat shock (Figure 8D).

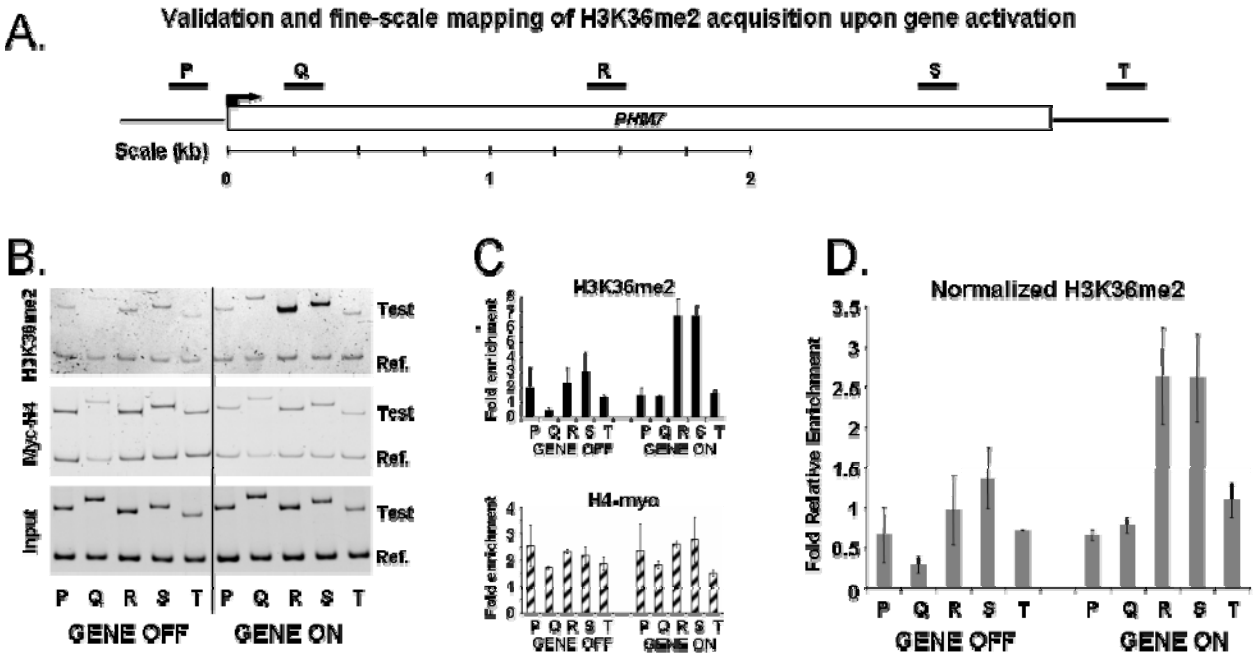


Figure 8 Validation and fine-scale mapping of H3K36me2 acquisition upon gene activation.

(A) Schematic of the *PHM7* locus, showing the locations of primers used to interrogate three of the independent H3K36me2 ChIP-chips shown in Fig. 2.5B, before and after heat shock. Primer sets are listed in Table 1. (B) Polyacrylamide gel analysis of PCR products generated by the primer sets shown in panel A (Test). The reference product (Ref.) is the same as that used for Fig. 2.2. (C) Quantitation of the gel shown in panel B. Graphs show the average enrichments detected by the PCR primer sets following H3K36me2 and histone H4-myc ChIPs, both before and after heat shock. The value plotted for each fragment was calculated as described in the legend for Fig. 2.2. (D) H3K36me2 enrichment values before and after heat shock after normalization with general nucleosome occupancy levels, calculated as described in the legend for Fig. 2.2. H3K36me2 levels increase after heat shock. Note that the increase is confined primarily to areas R, S, and T, which represent the center and 3' end of the gene, but not areas P and Q, which represent, respectively, the promoter and coding sequences at the 5' end. These experiments were performed by Yoichiro Shibata. Figure courtesy of Yoichiro Shibata.

A positive correlation between gene length and measured H3K36me2 ChIP enrichment suggests that H3K36 dimethylation is initiated at a fixed distance from the start of transcription. Previous studies at selected genes, including *ADH1*, *PYK1*, *PMA1*, and *SCC2*, have shown that dimethylation of H3K36 is initiated after transcriptional initiation, concomitant with association of Set2p with the elongating polymerase (Bannister et al., 2005; Kizer et al., 2005; Krogan et al., 2003). If this mechanism operates genome-wide, and if the interval between transcriptional initiation and Set2p association is constant regardless of gene length, longer genes will appear to be enriched by my H3K36me2 ChIPs to a greater extent than shorter genes. The reason for this predicted relationship is illustrated in Figure 9A and the corresponding legend and is a consequence of the fact that the DNA on my microarrays covers each ORF from the start codon to the stop codon, regardless of length.

To test this prediction, I plotted enrichment in H3K36me2 ChIPs against ORF length (Figure 9B). This analysis showed that longer ORFs appeared to be more efficiently enriched in H3K36me2 ChIPs, consistent with the prediction made by the hypothesis described above (Figure 9B). The "leveling off" observed at ~2,000 bp is also a predicted feature of a mark that begins at a set distance from the transcriptional start, since the proportion of the gene that is not modified becomes smaller with increasing length. No such relationship between ORF length and enrichment is observed with H3K36me2 ChIPs performed from *set2*Δ extracts (data not shown). I did observe a weak relationship between ORF length and apparent bulk nucleosome occupancy for the H4-myc ChIPs performed in this study (the effect was even less pronounced in reference (Lee et al., 2004). This may be due to nucleosome loss very near to the site of transcriptional initiation, which would be predicted to have this effect. In any case, the magnitude of the

relationship between size and length was much stronger for the H3K36me2 ChIPs, suggesting a defined boundary for the initiation of H3K36me2 inside the ORFs genome-wide. This conclusion is further supported by the H3K36me2 ChIP profile across 16 kb of chromosome XII (Figure 4D). For example, primer set A is situated ~154 bp from the *LEU3* transcription start site and does not detect significant enrichment, whereas chromatin represented by primer set E is situated ~400 bp after the start site of *SST2* and is highly enriched. Likewise, for the relatively long *FMP27* gene, primer set K at the 5' end of the coding region does not report enrichment, but the downstream primer sets L, M, and N show steady enrichment of the chromatin at the 3' end of the coding region.

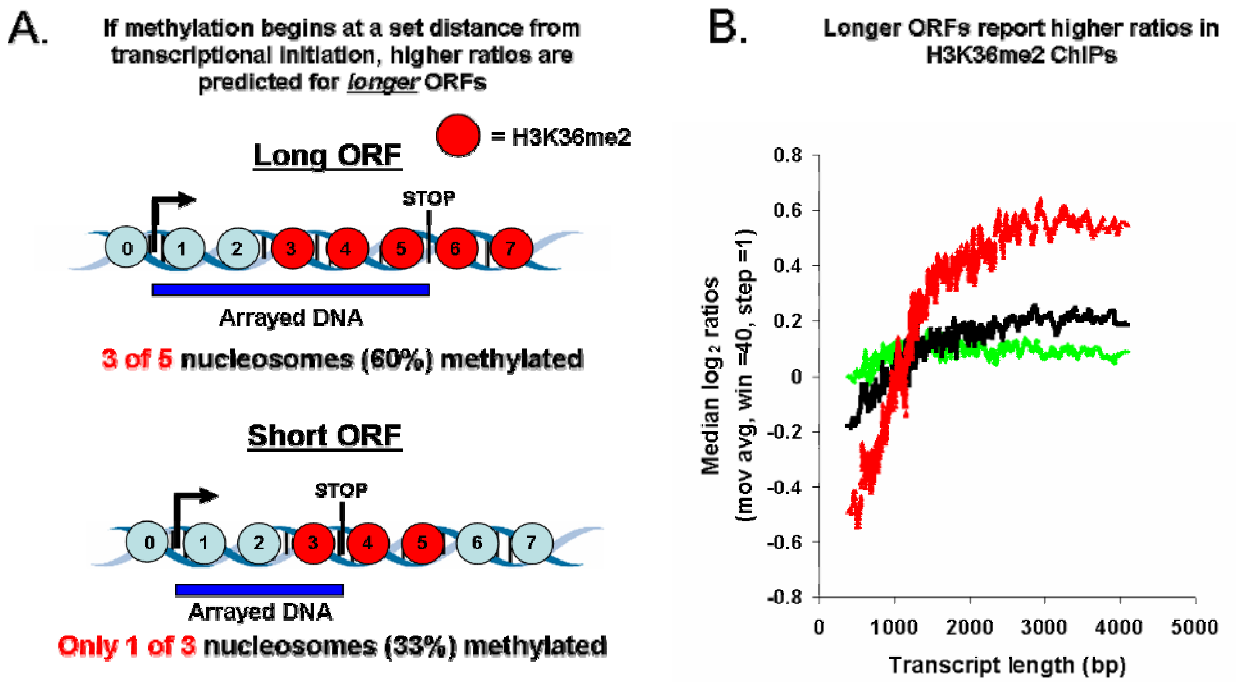


Figure 9 Evidence that H3K36me2 begins at a set distance from the initiation of transcription genome-wide.

(A) The hypothesis that H3K36me2 begins at a determined and consistent distance from transcriptional initiation predicts that higher ratios will be reported for longer ORFs. Note that "input chromatin" is used as a reference for these experiments, and the null expectation is that the raw signal intensities in both channels will be proportional to ORF length, resulting in neutral ratios for all ORFs. The prediction of higher ratios for longer ORFs is based entirely on the relative proportion of the arrayed element that is dimethylated, not the absolute length of the arrayed element or genomic feature per se. For example, if the entire ORF were modified, or if the distance at which the modification began from transcriptional initiation were proportional to ORF length, equal ratios would be obtained for long and short ORFs. Blue circles, nucleosomes not dimethylated at H3K36. (B) ORFs were sorted by length, and moving averages (mov avg) (window [win] = 40, step = 1) of their ratios of enrichment in H3K36me2 ChIPs (not normalized to H4-myc) and H4-myc ChIPs (from this study and reference (Lee et al., 2004) were plotted.

In the course of this analysis, I noted that short genes, on the whole, tend to be more frequently transcribed than long ones. Therefore, I wondered whether the correlations between length and ratio reported here confounded the conclusions presented in Figure 5. To test this possibility, only ORFs greater than 1,000 bp in length, which do not show a relationship between length and

transcription frequency, were used in the same analysis shown in Figure 5. The resulting plot was indistinguishable from the one presented.

Evidence that H3K36me2 ends upon termination of transcription. The results presented thus far provide evidence that dimethylation of H3K36 is restricted to transcribed genomic regions. A corollary to that hypothesis is that H3K36me2 ends upon transcriptional termination. This hypothesis predicts that the smaller the interval between two convergently transcribed genes, the higher the measured ratio of enrichment in my H3K36me2 ChIPs. This is because these shorter regions are likely to be entirely transcribed, while as the distance between the two upstream genes grows, it becomes progressively less likely that the entire intergenic region will be transcribed. This would result in unmodified nucleosomes toward the center of the fragment, resulting in lower ratios (illustrated in Figure 10A). Note that this prediction of the relationship between size and enrichment is the opposite of the previously described scenario for ORFs.

As predicted, I find that these shorter nonpromoter regions appear more highly enriched than longer nonpromoters in my H3K36me2 ChIPs (Figure 10B). In contrast, the sizes of the regulatory regions upstream of two divergently transcribed genes, which themselves are not transcribed, show no relationship to the degrees of ChIP enrichment (Figure 10C). Single promoters are expected to be partially transcribed, since they contain the 3' UTR of an upstream gene. As predicted, single promoters exhibit an inverse relationship (weaker than that observed for nonpromoters) between size and reported H3K36me2 ratio value (Figure 10D). No such relationship was observed when ChIPs were performed with histone H3 antibodies or in H4-myc strains or when H3K36me2 ChIPs were performed from *set2*Δ extracts (data not shown).

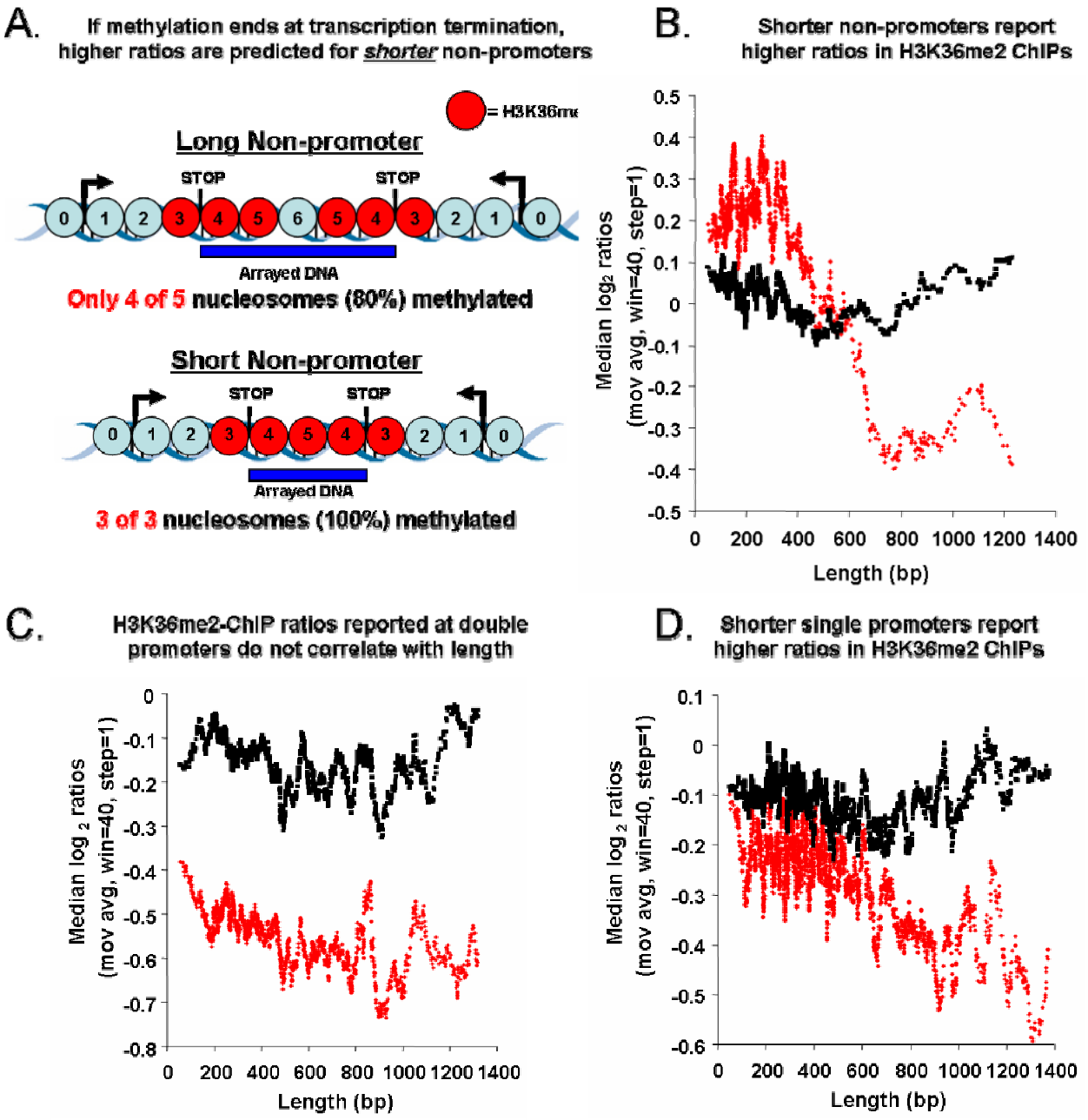


Figure 10 Evidence that H3K36me2 ends at transcriptional termination genome-wide.

(A) The hypothesis that H3K36 dimethylation ends upon transcriptional termination predicts that higher ratios will be reported for shorter nonpromoters (see the text for details). Blue circles, nucleosomes not dimethylated at H3K36. (B) Nonpromoters were sorted by length, and moving averages (mov avg) (window [win] = 40, step = 1) of their degrees of enrichment expressed as percentile ranks in H3K36me2 ChIPs (not normalized to H4-myc) and H4-myc ChIPs were plotted. (C) Same as panel B but for double promoters. (D) Same as panel B but for single promoters.

H3K36me2 is lacking in transcriptionally silent chromatin and in chromatin transcribed by RNA pol-III. To explore the possibility of other mechanisms of H3K36me2, I examined chromatin at telomeres and mating-type loci, two types of loci that are generally transcriptionally silent but serve specialized genomic functions. Both regions exhibit high nucleosome occupancy but are lacking in H3K36me2 (Figure 11). I also asked whether H3K36 dimethylation was specific to chromatin transcribed by RNA pol-II or whether other polymerases might support cotranscriptional modification. Due to the repetitive nature of the RNA pol-I transcribed rRNA genes, we were unable to make conclusions regarding H3K36me2 levels at these loci. However, I examined the RNA pol-III transcribed tRNA loci and found that these regions were not enriched by my H3K36me2 ChIPs (Figure 11). Although general nucleosome occupancy was also very low in regions transcribed by RNA pol-III, these results are consistent with the lack of evidence linking Set2p to RNA pol-III. Therefore, Set2p's function in dimethylation of H3K36 appears to be mediated exclusively through its association with RNA pol-II.

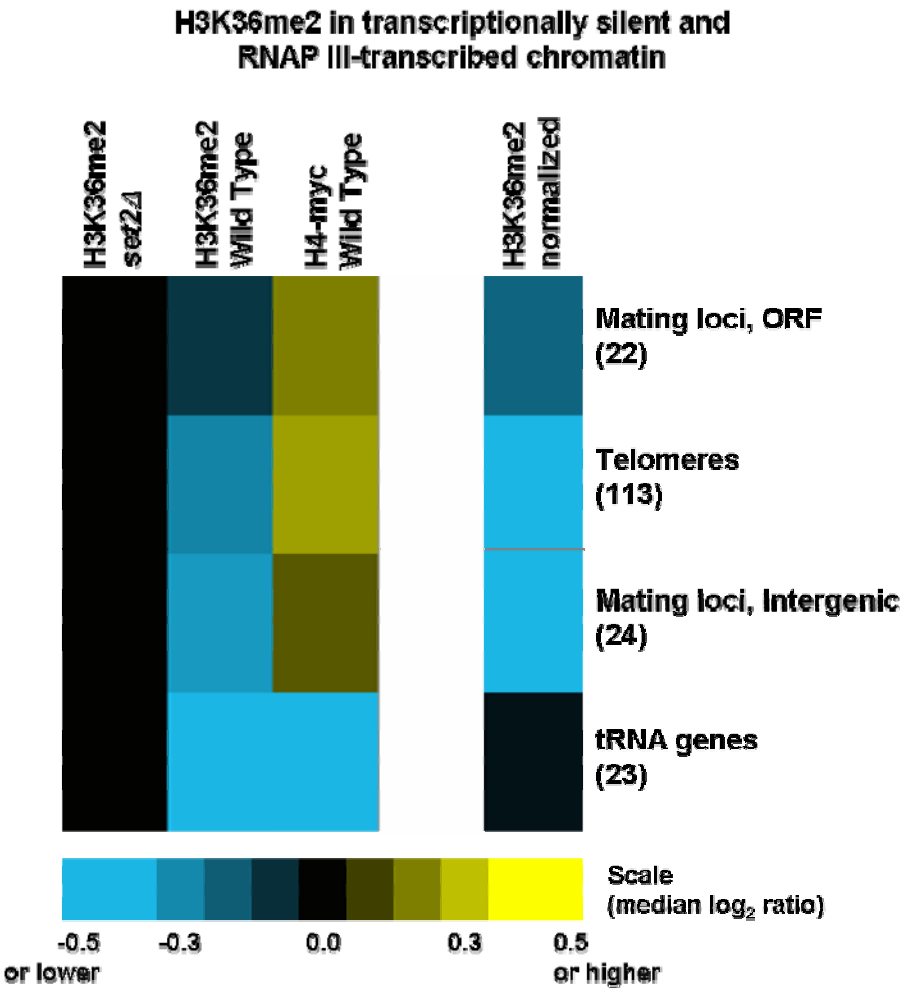


Figure 11 H3K36me2 levels are very low or absent in transcriptionally silent and RNA pol-III transcribed chromatin.

Colors (scale at bottom) represent the medians of reported ChIP ratios, as described in the legend for Fig. 2.1B.

DISCUSSION

While previous studies had examined the distribution and behavior of *S. cerevisiae* H3K36me2 at a few loci, it was not known whether the reported characteristics were representative of other genes and genomic regions. The genome-wide analyses of H3K36me2 distribution and dynamics

reported here allow us to unambiguously identify the core properties of this epigenetic mark in *S. cerevisiae*. The most important of these emergent properties are as follows. (i) H3K36me2 is scarce or absent in upstream gene regulatory regions, telomeres, mating loci, and regions transcribed by RNA pol-III. This provides evidence that the enzymatic activity of Set2p is firmly restricted to its association with RNA pol-II. (ii) The degree of H3K36me2 within ORFs correlates with the "on" or "off" state of transcription, but among genes that are measurably transcribed, the degree of modification does not correlate with the frequency of transcription. This provides evidence that H3K36me2 occurs chiefly in the initial instance of gene transcription, with subsequent rounds playing at most a maintenance role. (iii) In support of the previous point, newly transcribed genes that had been transcriptionally dormant acquire the H3K36me2 mark but at levels that do not correlate with the degree of induction. (iv) Set2p enzymatic activity begins in ORFs at a fairly constant distance from the initiation of transcription, which does not vary with gene length. (v) Nucleosomes in noncoding regions immediately downstream of transcribed genes are as highly dimethylated at H3K36 as nucleosomes in ORFs. This result provides evidence that once Set2p is associated with RNA pol-II, Set2p continues to be active and remains associated with RNA pol-II throughout the rest of the transcript length. (vi) With increasing distance downstream of the stop codon, H3K36me2 levels decrease. This result provides evidence that Set2p histone methyltransferase activity is terminated along with the termination of transcription. (vii) Unlike the nucleosome depletion that is normally observed in the ORFs of highly transcribed chromatin, H3K36 dimethylated nucleosomes appear to be refractory to depletion, suggesting preferential retention of H3K36 methylated nucleosomes on DNA. However, nucleosome dynamics in a *set2* deletion strain appeared to be normal, indicating that while H3K36me2 may be an indicator of stability, it is not

likely to be a required factor. Although there may be some variation from locus to locus, these core properties shed light on important questions surrounding the function of the Set2p enzyme and H3K36me2 itself.

Challenges in determining the global distribution of chromatin modifications. Before discussion of the biological function of H3K36me2, it is worth mentioning some of the challenges that are inherent to any experiment that aims to determine the distribution of a histone modification genome-wide. In this study, we used ChIP to specifically enrich for genomic regions that contain H3K36me2 nucleosomes and then interpreted the efficiency of DNA recovery at each locus to reflect the relative amount of H3K36 dimethylation at each locus. Using this approach, several factors could create nonbiological variation in results, including the effects of fixation, epitope accessibility, antibody specificity, microarray content, and underlying bulk nucleosome occupancy. These challenges have been discussed at length in recent reviews (Buck and Lieb, 2004; Hanlon and Lieb, 2004; Loden and van Steensel, 2005; van Steensel and Henikoff, 2003).

This study includes important advances in addressing some of these issues. First, I thoroughly demonstrated the specificity of my H3K36me2 antibody by dot blot against H3K36me0, H3K36me1, and H3K36me3 peptides (Figure 3A), Western blots derived from whole-cell and nuclear extracts and control ChIPs in *set2*Δ strains (Figure 3B). Second, I used DNA microarrays that cover the entire genome on a single slide. This represents a significant improvement over many published studies that used arrays representing only the ORFs or only noncoding intergenic regions or that split ChIP samples and hybridized them independently to separate arrays representing only the ORFs or the intergenic regions. Use of a whole-genome array was essential

to most of the conclusions presented here (Hanlon and Lieb, 2004). Third, the H3K36me2 data have been normalized to bulk nucleosome occupancy, using data from H3 or H4-myc ChIPs that were performed in parallel from the same extract. This is important because recent studies have shown that nucleosome occupancy throughout the yeast genome is heterogeneous (Bernstein et al., 2004; Lee et al., 2004), and if left unaccounted for, misleading patterns could emerge. To our knowledge, this is the first instance of modified-nucleosome ChIP data being normalized to apparent bulk nucleosome occupancy genome-wide. Finally, each of the ChIP-chip experiments were followed up with high-resolution PCR-based detection at individual loci (PCRs were performed by Yoichiro Shibata, our collaborator), which provided additional information and confirmed the conclusions drawn from the array results.

How is the intricate genomic pattern of histone H3K36me2 specified? The general mechanism of directing Set2p to specific genomic regions by piggybacking on elongating RNA pol-II through association with a doubly modified CTD (Ser2/Ser5) is entirely sufficient to explain the pattern of H3K36me2 we observed throughout the genome. This is an important conclusion because it indicates that Set2p modifies chromatin only when associated with RNA pol-II and not, for example, on soluble histone H3 prior to chromatin assembly. More specifically, the genome-wide analysis shows that H3K36me2 occurs at a determined distance from the initiation of transcription, regardless of ultimate transcript length (Figure 9). This result is consistent with PCR-based ChIP assays performed on single genes (Bannister et al., 2005; Kizer et al., 2005; Krogan et al., 2003; Schaft et al., 2003) and with locus-specific results presented here that imply H3K36me2 begins approximately two nucleosomes downstream of the start codon. In addition, the data indicate that Set2p chromatin-modifying activity stops upon

transcriptional termination (Figure 10). I observed no relationship between the presence of introns and H3K36me2 levels.

What is the biologically relevant function of H3K36me2? Our results indicate that levels of H3K36me2 are not correlated with the frequency of transcription but rather with the occurrence of transcription per se (Figure 5 and Figure 7). This result suggests that H3K36 methylation does not generally act as a "rheostat" for gene transcription. In *S. cerevisiae*, *set2*Δ strains are viable and, in many backgrounds, exhibit only mild phenotypes. So, what does H3K36 methylation do?

At individual loci, Set2p has been shown to act as a transcriptional repressor (Landry et al., 2003; Strahl et al., 2002). In one of these studies, Set2p caused repression of *GAL4* but not of other examined genes, and in the other, Set2p was artificially tethered to a promoter, which resulted in transcriptional repression. So while Set2p may act as a transcriptional repressor at individual genes or have the capacity to repress transcription if inappropriately tethered at promoters, a general role for repression of transcription at gene promoters is not consistent with the genomic pattern reported here.

Given Set2p's established interaction with elongating polymerase and the genomic pattern of H3K36me2 reported here, it is easy to envision a role for Set2p and H3K36me2 in transcriptional elongation. Several lines of evidence suggest that this is the case. Perhaps the most convincing is synthetic genetic array analysis, which revealed growth defects when a *set2*Δ mutant was combined with deletions of any of the five components of the Paf1 complex or of the transcription elongation factors Chd1p or Soh1p (Krogan et al., 2003). It has also been shown that deletion of genes encoding either of two components of the Paf1 complex, Rtf1p or Cdc73p,

resulted in a decrease in the recruitment of Set2p across the *PMA1* gene and abolished H3K36 dimethylation at that locus (Krogan et al., 2003). In addition to these findings, studies involving 6-azauracil help to confirm a role for Set2 as an elongation factor (Kizer et al., 2005; Krogan et al., 2003; Li et al., 2003). However, it is still not clear exactly how Set2p or H3K36me2 might participate in the process of elongation itself. It remains possible that Set2p's association with elongation is solely a mechanism to control the distribution of H3K36me2, rather than an indication of any direct participation in the transcription elongation process. In this case, the defects in elongation observed in the absence of H3K36me2 could be indirect consequences of a failure to recruit chromatin-modifying enzymes or other factors important for transcriptional elongation to ORFs.

An epigenetic mark to distinguish regulatory and nonregulatory chromatin genome-wide. A mark such as H3K36me2 could also function as a "molecular memory" of transcription patterns that are specified at only one point during development or the life cycle but must be maintained afterwards. This concept of transcriptional memory is similar to what has been proposed for *S. cerevisiae* H3K4me3, which remains stable on chromatin long after the transcription of chromatin at that locus has ceased (Ng et al., 2003b). The putative "memory" role of H3K36 methylation, which appears to be similarly stable, could be accomplished by the ability of this mark to physically affect the chromatin fiber or, more likely, through the recruitment of other remodeling factors that alter chromatin structure (Nakayama et al., 2001; Rea et al., 2000). For example, it has been recently observed that histone H3 and H4 acetylation is generally lower in coding regions than in promoters and that this global acetylation pattern is regulated by the protein Eaf3p (Reid et al., 2004). Eaf3 is a subunit of both the NuA4 histone acetylase complex

and the Rpd3 histone deacetylase complex (Deckert and Struhl, 2002; Gavin et al., 2002). Reid et al. proposed that "Eaf3 might recognize some feature of chromatin (e.g., nucleosome conformation or nonhistone protein) that is distinct between promoters and coding regions" (Reid et al., 2004). H3K36 methylation could be just such a distinguishing feature of coding and noncoding chromatin.

One intriguing possibility along these lines is that Set2p mediates H3K36me2 to create a stable epigenetic mark that generally serves to distinguish regulatory and nonregulatory chromatin genome-wide. What function might such a distinction serve? As described in the introduction, coding and noncoding chromatin exhibit several biologically important differences whose underlying physical basis remains unexplained. Higher nucleosome occupancy in the body of genes may serve to prevent nonproductive transcription factor-DNA interactions by occluding binding sites that occur in coding regions. Conversely, nucleosomes in promoter regions may be more prone to low occupancy or disassembly, thereby exposing binding sites and directing transcription factors to appropriate targets (Sekinger et al., 2005b). These two tendencies, acting in concert, would have the effect of reducing the "sequence space" that must be searched by any given factor before an appropriate target is found.

In *S. cerevisiae*, about 85% of genes are transcribed at detectable levels during mitotic growth (Holstege et al., 1998; Hurowitz and Brown, 2003), meaning that H3K36me2 distinguishes regulatory and nonregulatory chromatin throughout most of the genome. In actively transcribed chromatin domains, upstream regulatory sequences are clearly distinguished by their lack of the H3K36me2 mark. Therefore, the H3K36me2 mark, which is conserved throughout eukaryotic evolution, represents the first physical mark that has been shown to distinguish regulatory

sequences from coding and nonregulatory intergenic sequences genome-wide. A transcription-coupled mark that does not correlate with transcription rate and is correlated with stabilized nucleosomes on transcribed chromatin, both properties of H3K36me2 described here, may be a general feature of eukaryotic chromatin that contributes to the mechanism of context-dependent targeting of DNA-associated proteins.

Table 1 PCR primers used in Figure 4 and Figure 8.

Set	Forward Primer (5′ to 3′)	Reverse Primer (5′ to 3′)
A	GGAGCCATGCACTAAGTGTG	CGTCTCAAAAGCACTCTTCCT
B	TAAGGCATCCAGAAAAGAAAGC	CAGAGGTTCATTGACAGCTTCAG
C	CTGCGGTGCTACAGGAGAGGTAA	CGTGAGCGCTTACGAATCTTC
D	CGTTAGGTGAACCACCTGAAGTG	CCTGAGGACCTTCAAAGTTGC
E	CACCTAAGGGAGTTGCTGTTTTAC	GGCAAGGTAATGGGTCATCTG
F	GAAGAAGCCATGAAGTATGCGAAG	GGACGATGAGAAGTAAAGACAGAG
G	CAGTCTTACAACTCGCTTGTCAC	GTGGAAGACGCTGGTTTTTC
H	GTGTTAGCAGCCCAAAGGTG	CTGAAACTGCACGGTCGATAC
I	CCACGTTCGTATTTATATCTATATTCG	CCGTTCTTGCTGACATGGTTTTC
J	CCTATTTTTTTGGCTTGTAACTCAAAC	CTGTCCTTTGTAATAGCTACCGTTC
K	CCGATTAATGTTCTACTGTACAAGTGG	CATTAGTAGATGATTTCTTGGAATCA
L	GGTGCTAGGTCGGTTTTTATG	CGGCAATACTGGACTTATGGTTAG
M	CGCAGATGTCTCACCATTATTC	GATGTCCACGTCCTGGCATAC
N	GCCGTTGCAAAGCTAAACTTTC	GTGGTCTCAACGTTTTTAACAACC
O	GTAGCATCATACATATCAATATGTTACCA	CATGGTCTCTTGGCCCAGTTG
P	CGTTAATCTCCCTGAAATACTTTAAATAC	CGCTGCTTTGAATATGATGTGC
Q	GGAGGAAGAGAGAACGGAACC	GTTGGTAACATTTGAAAATGATAGCAG
R	CTTCCTGATGGGTGTCATTACT	GTCAATAATCGAGTCAACGGTG
S	CGATGTTGAGGCATTGAGAAATAAG	CTCATCCTGACACTGTCGAAAG
T	CCGCATATCTGTGAGCTTGG	GATAGACACTCAACCTATAGGAACTAGTTTC

Chapter III

Section I
ACETYLATED HISTONE H3 LYSINE 36 IS LOCALIZED TO PROMOTERS OF RNA POLYMERASE II TRANSCRIBED GENES GENOME WIDE

This work was published as a part of a collaborative project under the title:
Identification of histone H3 lysine 36 acetylation as a highly conserved histone modification

Stephanie A. Morris, Bhargavi Rao, Benjamin A. Garcia, Sandra B. Hake, Robert L. Diaz, Jeffrey Shabanowitz, Donald F. Hunt, C. David Allis, Jason D. Lieb, and Brian D. Strahl
*Journal of Biological Chemistr*y, Mar 9;282(10):7632-40

INTRODUCTION

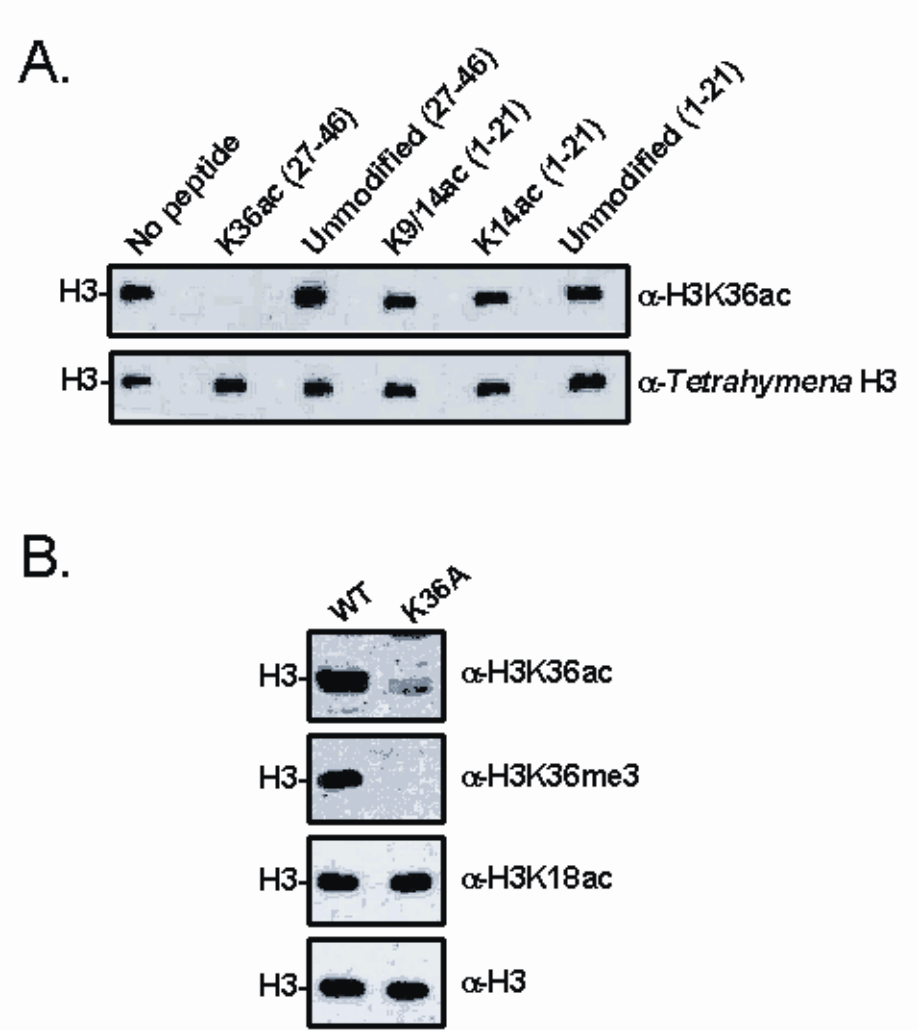
Histone acetylation is one of the most extensively studied histone modifications. Histone acetylation has traditionally been known to be involved in transcriptional activation. Acetylation has been shown to alter the secondary structure of the histone tail and weaken histone-DNA interactions resulting in the opening of the chromatin structure to allow transcription. Specific acetylation patterns may also function to recruit other factors that regulate the chromatin structure. Role of histone acetylation in nucleosome displacement was demonstrated from studies that showed histones are acetylated prior to their removal from the promoter of the *PHO5* gene after it has been activated (Reinke and Horz, 2003). A direct link between histone removal and acetylation comes from the fact that the SWI/SNF complex is recruited to acetylated nucleosomes and is responsible for the loss of histones (Reinke and Horz, 2003; Steger and Workman, 1996).

Acetylation is highly dynamic and has been linked to cellular processes such as transcriptional activation, DNA repair, as well as chromatin assembly (Masumoto et al., 2005; Wang et al., 1997; Ye et al., 2005). Genome wide studies on the histone acetyltransferases (HATs) Gcn5p and Esa1p show that they are recruited to the promoters of inducible genes (Robert et al., 2004). These studies show that DNA binding transcription factors recruit these complexes to nucleosomes near UAS elements near protein coding genes. Acetylation further stimulates trimethylation of histone H3K4 in the 5′ region of coding regions (Govind et al., 2007). Studies have shown that Gcn5p in isolation can only target H3K14 (Grant et al., 1999; Kuo et al., 1996). However, in its native SAGA complex this enzyme has an expanded substrate range on H3 that includes H3K9, K18 and K23 (Grant et al., 1999). In mammals and fission yeast that H3K9 methylation serves as a binding site for the recruitment of the chromodomain protein HP1, initiating the formation of heterochromatin (Bannister et al., 2001; Nakayama et al., 2001). Until recently, H3K9 was the only known residue found to be acetylated or methylated.

BACKGROUND

In collaboration with Stephanie Morris in Dr. Brian Strahl's lab I sought to determine if H3K36 in yeast was acetylated. H3K36 as shown by our studies and others is known to be methylated. H3K36me occurs at the 3' region of coding genes. Using mass spectrometric analysis, our collaborators analyzed histone H3 in *Tetrahymena*. The results identified a novel acetylation event at H3K36 that is conserved between *Tetrahymena* and yeast.

An antibody specific to acetylated H3K36 was raised against a synthetic peptide acetylated at H3K36, and *Tetrahymena* and yeast histones were analyzed by immunoblot analysis (Fig. 3.1A). In vivo analysis in yeast showed that the antibody specifically recognizes acetylated H3K36 and does not cross react with H3K36me and acetylated H3K18 (Figure 12B).



(Figure reproduced from Morris *et al*, 2005. J Biol Chem. 2007 Mar 9;282(10):7632-40)

Figure 12 Detection of H3K36 acetylation in *Tetrahymena* and yeast using a specific antiserum.

(A) An antibody specific to H3K36 acetylation recognizes *Tetrahymena* H3. RP-HPLC *Tetrahymena* H3 was loaded onto adjacent lanes and resolved on a 15% SDS-PAGE gel. Following transfer to a PVDF membrane, each lane was separated and probed with an α-H3K36 acetyl antiserum (α-H3K36ac) that was preincubated with different unmodified or modified H3 synthetic peptides as indicated. The same blots were stripped and reprobed with an antibody specific for *Tetrahymena* H3 (α-*Tetrahymena* H3) as a loading control. (B) The α-H3K36ac antibody specifically recognizes H3K36 acetylation in yeast. Acid-extracted histones prepared from a wild-type or H3K36 point mutant yeast strain (K36A) were resolved on a 15% SDS-PAGE gel, transferred to a PVDF membrane, and probed for H3K36ac. An antibody specific for the C terminus of H3 (α-H3) was used as a loading control. Antibodies specific for H3K18 acetylation (α-H3K18ac) and H3K36 trimethylation (α-H3K36me3) were used as additional controls. The western blots were performed by Stephanie Morris. Figure is reproduced from Morris *et al*, *JBC*, 2007.

Our collaborators then sought to identify the HAT(s) responsible for acetylation of H3K36. Using TAP-purified Gcn5p (subunit of SAGA complex), they incubated the complex with either unmodified or modified H3 synthetic peptides along with unlabeled acetyl-CoA (acetyl donor) in a HAT assay. The products were then electrophoresed on SDS-PAGE gels and analyzed by immunoblot with the α-H3K36ac antibody. After incubation with purified SAGA, a previously unmodified H3 27-46 amino acid peptide was recognized by the H3K36ac-specific antibody. When assays were performed using a radiolabeled form of acetyl-CoA, SAGA was readily able to acetylate an N-terminal H3 peptide (residues 1-21 containing H3K9 and H3K14) as well as the unmodified 27-46 peptide, but showed no activity toward a matched 27-46 peptide acetylated at H3K36 (Figure 12A). Deletion of *GCN5* resulted in a loss of acetylated H3K36.

After the presence of the modification had been established, the next important question was determining the *in vivo* function of acetylated H3K36. Using the ChIP assay and then hybridizing the isolated DNA on whole genome yeast DNA microarrays (ChIP-chip), I wanted to investigate whether this modification was localized to certain genomic regions, whether it is preferentially localized to highly transcribed regions and whether it was present in regions not transcribed by RNA pol-II. In order to further our understanding of the role of acetylation of

various histone residues and the interplay between acetylation methylation, we directly compared patterns of H3K36 acetylation and methylation.

MATERIALS AND METHODS

DNA microarray (ChIP-chip) analyses. ChIP assays were performed as previously described (Xiao et al., 2003). Antibodies and amounts used in the immunoprecipitations (IPs) are as follows: α-H3K36ac (2 μL/IP), α-H3K36me2 (3 μL/IP, Upstate Biotechnology, catalog # 07-274), and α-H3K9/14ac (3 μL/IP, Upstate Biotechnology, catalog # 06-599). Following DNA recovery, the ChIP-enriched DNAs were amplified as described (Bohlander et al., 1992). Briefly, two initial rounds of DNA synthesis with T7 DNA polymerase using primer 1 (5'-GTTTCCCAGTCACGATCNNNNNNNNN-3') was followed by 25 cycles of PCR with primer 2 (5'-GTTTCCCAGTCACGATC-3'). Cy3-dUTP or Cy5-dUTP were then incorporated directly with an additional 25 cycles of PCR using primer 2. Direct microarray hybridizations of H3K36ac ChIP vs. H3K36me2 ChIP or H3K9/K14ac ChIP vs. H3K36me2 ChIP were performed using standard procedures (Iyer et al., 2001). This method allowed for the direct comparison between the histone modification patterns, which showed that H3K36ac enrichment was preferentially in the promoter regions of genes while H3K36me2 enrichment was preferentially in the transcribed regions. Additional control experiments in which the reference DNA employed was from H3 ChIPs (for standard nucleosome occupancy normalization), H3K36ac ChIPs from a H3K36 point mutation strain (K36A), or genomic DNA also demonstrated the H3K36ac enrichment to be promoter specific (data not shown). Following hybridizations, the

arrays were scanned with a GenePix 4000 scanner and data was extracted with Genepix 5.0 software. Data were normalized such that the median $\log_2$ ratio value for all quality.

RESULTS

H3K36 acetylation is preferentially enriched in the promoters of RNA pol-II transcribed genes genome-wide. I used a ChIP-chip approach to determine the genomic distribution of H3K36ac and how it compared to the distribution of methylation found at H3K36. The α-H3K36ac-specific antibody was used in ChIP reactions from yeast whole cell extracts. The ChIP reactions were performed by my collaborator, Stephanie Morris. Enriched genomic DNA fragments were treated with RNase, amplified, and labeled fluorescently. DNA enriched from H3K36me2 ChIPs was prepared in a similar manner, and both samples were hybridized on the same array. Three independent sets of ChIPs were performed. Using this method, I directly compared the patterns of H3K36ac and H3K36me2 using arrays that tiled continuously over the entire genome at a resolution of ~1 kb. I observed a preferential enrichment of H3K36ac at regulatory regions (bidirectional and unidirectional promoters) relative to coding regions (ORFs) and 3′ UTRs (Figure 13A). Significantly, the H3K36ac pattern was found to be inversely related to that of H3K36me2, which occurs preferentially in the coding region and 3′ UTR of genes.

I then compared the location of this modification to that of other well-characterized acetylation sites on H3; namely H3K9 and H3K14 acetylation (H3K9/14ac). Using the same extracts, ChIPs were performed with an antibody that recognizes diacetylated H3K9/14, and the enriched DNA was amplified, labeled, and hybridized to DNA microarrays as described above. These experiments revealed that H3K9/K14ac was localized to promoters in a pattern strikingly similar

to the pattern I observed for H3K36ac (data not shown and see Figure 13B). These data were also consistent with the H3K9/14 acetylation results obtained by others (Liu et al., 2005; Pokholok et al., 2005). I further examined the distribution of H3K36ac and H3K9/K14ac using a high-resolution microarray containing probes that covered all of chromosome III at a resolution of 200 bp (and at 100 bp resolution for 1/3 of the chromosome). For both H3K36ac and H3K9/K14ac, the level of acetylation enrichment drops sharply as a function of distance from translational initiation site (Figure 13B). These data are fully consistent with our analysis using lower-resolution arrays and confirm that these acetyl marks occur preferentially upstream of coding regions.

I next asked whether H3K36ac associates with genomic regions other than those characteristic of RNA pol-II promoters. I found that genomic regions which were transcriptionally silent or regions not transcribed by RNA pol-II, including centromeres, telomeres, and mating type loci, have low levels of H3K36ac and H3K9/K14ac (Figure 13C). These results indicate that H3K36ac is associated strictly with active RNA pol-II regulatory sequences, and suggest that like H3K36me, H3K36ac may function in RNA pol-II mediated gene transcription.

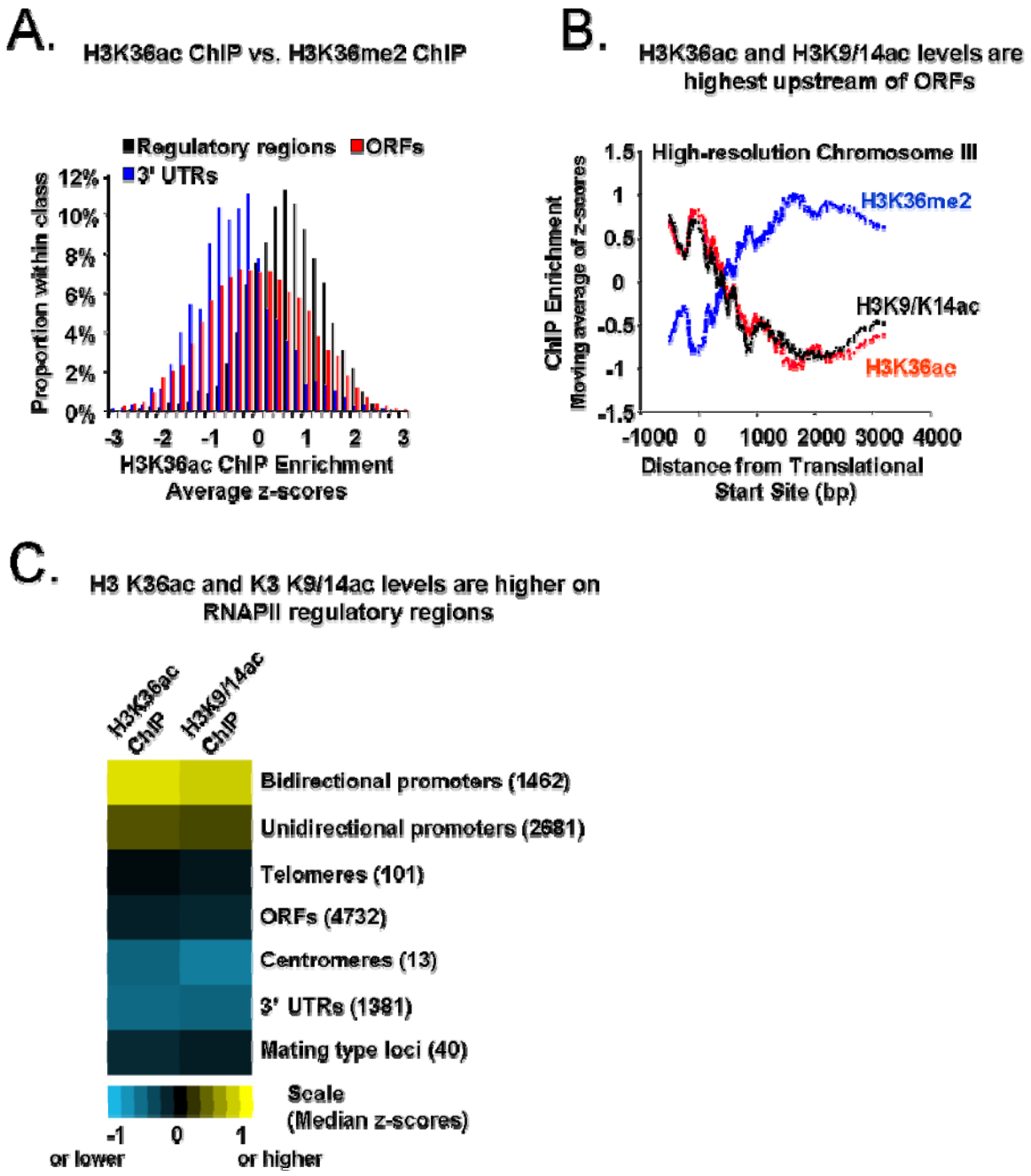


Figure 13 H3K36 acetylation is localized predominantly to the promoters of RNA pol-II transcribed genes genome-wide.

(A) The distribution of average *z*-scores (units are standard deviation from the mean) for 5' regulatory regions (black), open reading frames (red), and 3'-untranslated regions (blue) derived from ChIP-chip experiments in which H3K36ac ChIPs were compared directly to H3K36me2 ChIPs. Thus, the H3K36ac and H3K36me2 ratios shown here are inversely related. Similar promoter enrichment

results for H3K36ac were obtained when the H3K36ac ChIPs were compared with a genomic DNA reference, or references composed of histone H3 ChIPs. (B) A moving average (mov avg) (window [win] = 40, step = 1) of average *z*-scores from three independent experiments comparing H3K36ac ChIPs (red), H3K9/14ac ChIPs (black), and H3K36me2 ChIPs (blue) on a high resolution DNA microarray covering all of chromosome III. ChIP enrichment is plotted as a function of the distance from the translational start site among genes greater than 1 kb in length. (C) H3K36ac distribution genomewide. Colors (scale at bottom) represent the median of all *z*-scores recorded from all arrayed elements in the indicated functional class (labeled on the right; number of elements is indicated in parentheses). Data were derived from three independent replicates.

Section II

ACETYLATED HISTONE H3 LYSINE 36 IS LOCALIZED TO PROMOTERS OF RNA POLYMERASE II TRANSCRIBED GENES GENOME WIDE

This work was published as a part of a collaborative project under the title:
The RNA polymerase II kinase Ctk1 regulates positioning of a 5' histone methylation boundary along genes

Tiaojiang Xiao, Yoichiro Shibata, Bhargavi Rao, R. Nicholas Laribee, Rose O'Rourke, Michael J. Buck, Jack F. Greenblatt, Nevan J. Krogan, Jason D. Lieb, and Brian D. Strahl
Molecular and Cellular Biology, 2007 Jan;27(2):721-31

INTRODUCTION

In *Saccharomyces cerevisiae*, lysines 4, 36 and 79 are methylated by Set1p, Set2p and Dot1p respectively (Boa et al., 2003; Briggs et al., 2001; Lacoste et al., 2002; Roguev et al., 2001; Strahl et al., 2002). Set2p has been shown to methylate H3K36 in the 3' end of coding region and this methylation event requires association of Set2p with Ser-2) phosphorylated CTD of RNA pol-II. This phosphorylation event is mediated by Ctk1 enzyme complex. Set1p, on the other hand, has been found to be predominantly localized to 5' end of the genes and presence of H3K4me3 corresponds to Set1p occupancy (Ng et al., 2003b). Set1p associates with ser-5 phosphorylated CTD of RNA pol-II (Ng et al., 2003b). Kin28, the ser-5 CTD kinase, mediates this phosphorylation event. Set1p associates with RNA pol-II through the Paf1 complex consisting of Paf1p, Ctr9p, Cdc73p, Rtf1p and Leo1p (Krogan et al., 2003). Set1p is the sole HMT in yeast that is responsible for methylating all three mono-, di- and trimethyl forms of H3K4.

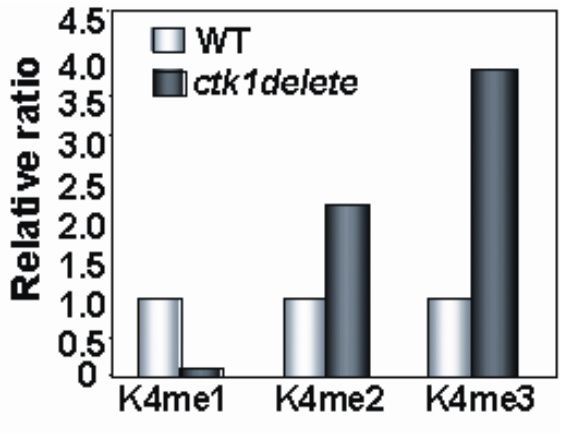
H3K4me3 and Set1p has been shown to be localized specifically to 5' regions of actively transcribing genes (Santos-Rosa et al., 2002), it is speculated that Set1p associates with RNA pol-II early in the elongation phase. H3K4me2, on the other hand, was found to be present mostly in the middle of the genes andH3K4me1 was enriched at the end of the genes (Pokholok et al., 2005). H3K4me has been shown to recruit chromatin remodeling complexes. In yeast, H3K4 methylation recruits both the Isw1 and Chd1 ATPases (Pray-Grant et al., 2005; Santos-Rosa et al., 2003). More recently studies have shown a plant homeodomain (PHD) finger containing proteins that specifically bind to methylated lysine and regulate gene expression (Li et al., 2006; Pena et al., 2006; Shi et al., 2006; Wysocka et al., 2006).

Although H3K4 mono-, di- and trimethylation is carried out by a single enzyme, there is mounting evidence that they are regulated by different pathways. Different proteins can bind histone modifications with very high specificity thereby distinguishing the same modification at different residues.

BACKGROUND

In an attempt to identify pathways that selectively regulate distinct H3K4 methylation states, my colleague in Dr. Brian Strahl's lab, Dr. Tiaojiang Xiao, conducted a screen by carrying out western analyses of H3K4me1 using whole cell extracts (WCEs) derived from 384 individual strains of yeast containing single deletions of genes known or predicted to function in some aspect of chromatin function or transcription regulation. Among other factors, the screen identified *CTK1*, *CTK2*, and *CTK3* members of the CTDK-1 kinase complex responsible for Ser-2 phosphorylation of CTD (Sterner et al., 1995), This factor had not been shown to regulate H3K4me1 previously. Analyses of other modifications of H3K4 in a *ctk1* deletion strain revealed

that while monomethylation was significantly reduced, di- and trimethylation levels were markedly increased (Figure 14) and a corresponding 5' spread of these methyl marks into body of transcribed genes in the absence of this kinase.



(Figure reproduced from Xiao *et al*, 2007 Mol Cell Biol. 2007 Jan;27(2):721-31)

Figure 14 Ctk1 kinase activity regulates H3K4 methylation.

Quantification of *ctk1Δ* effects on the relative proportion of H3 that was H3K4 mono-, di-, and trimethylated. Western blots were performed by Dr. T Xiao. Figure is reproduced from Xiao *et al*, *MCB*, 2007.

My collaborators showed that the spread of H3K4 methylation in the *ctk1*Δ strain was independent of Set2/K36 methylation. There was a modest increase in H3K4me3 in the absence of *SET2*. In contrast, *ctk1*Δ shows a dramatic spread of the modification into the 3' end of *FMP27*. These data imply Ctk1p regulates H3K4 methylation through a mechanism that is distinct from its regulation of Set2p and H3K36 methylation. *ctk1*Δ strain does not show increased levels of Set1p indicating that the increased levels of H3K4me3 was not due to increase in the protein levels. In order to explore the possibility of 'transcriptional stress' being the cause of increased H3K4me3, our collaborators looked at the H3K4me3 levels in *rpb1*ts and *kin28*ts mutants and in the presence of 6AU (a drug used to detect transcription elongation defects). By Western blot, they found that these treatments had no effect on global H3K4

methylation levels. These results indicated that Ctk1p acts to maintain H3K4 trimethylation at the 5' end of genes through a mechanism independent of unregulated RNA pol-II initiation within coding regions, observable elongation defects, or transcriptional stress.

Studies have shown that in a *set2*Δ strain there is aberrant transcription initiating from within the genes and with a corresponding increase in H3K4me3 (Carrozza et al., 2005). We therefore wanted to explore the possibility of increased H3K4me3 levels in a *ctk1*Δ strain was due to aberrant initiation occurring in genes. If this were to be the case then we would expect greater levels of aberrant transcription in *ctk1*Δ strain than in *set2*Δ strain. I sought to investigate this possibility comparing the levels of aberrant transcripts in *STE11* gene in *ctk1*Δ and *set2*Δ strains.

I also wanted to investigate the extent of the spread of H3K4me3 into the 3' end, whether Ctk1p regulated the spread of this modification on the entire genome or if there were a particular class of genes that were affected. In order to address this question, I performed ChIP-chip studies comparing the distribution of H3K4me3 in wildtype and in *ctk1*Δ strain.

MATERIALS AND METHODS

Northern blot analyses Yeast Strains were grown at 30°C in YPD (1% yeast extract, 2% peptone, 2% dextrose) to an OD_{600} of 0.6-0.8. Total RNA was prepared by acid-phenol method as described (Xiao et al., 2003). For northern blotting, 30ug of each sample was subjected to electrophoresis in a formaldehyde-agarose gel, followed by transfer to a nylon membrane and crosslinked by UV irradiation. Hybridization was carried out in 6X SCP (2.5M NaCl, 0.7M

Na_2HPO_4, 0.02M EDTA), 1% Sarkosyl and 0.1 mg/ml of salmon sperm DNA. Probes (full-length STE11 and for FMP27 (+5051- +7346) were generated by PCR.

ChIP-chip analyses: DNA Amplification, Labeling, Array Hybridization and Data Processing. ChIP-enriched DNA and reference DNA in all experiments was amplified as described (Bohlander et al., 1992). Briefly, two initial rounds of DNA synthesis with T7 DNA polymerase using primer 1 (5'-GTTTCCCAGTCACGATCNNNNNNNNN-3') was followed by 25 cycles of PCR with primer 2 (5'-GTTTCCCAGTCACGATC-3'). Cy3-dUTP or Cy5-dUTP were then incorporated directly with an additional 25 cycles of PCR using primer 2. Microarray hybridizations were performed using standard procedures (Iyer et al., 2001). The arrays were scanned with a GenePix 4000 scanner and data was extracted with Genepix 5.0 software. Data were normalized such that the median $\log_2$ ratio value for all quality elements on each array equaled zero, and the median of pixel ratio values was retrieved for each spot. Only spots of high quality by visual inspection, with at least 50 pixels of quality data (regression $R^2 > 0.6$) were used for analysis. Arrayed elements that did not meet these criteria on at least half of the arrays were excluded from analysis. All data was log-transformed before further analysis. For ChIP-chip data analysis, the $\log_2$ ratio of each spot was transformed to a z-score using the formula, $z_x=(X-\mu)/\sigma$, where X is a retrieved spot value, μ is the mean of all retrieved spots from one array, and σ is the standard deviation of all retrieved spots from that same array. Z-scores from biological replicates were averaged. Same transformation was done for the histone H3 arrays. For normalization with the nucleosome occupancy data, for each probe the average z-score from the histone ChIP-chips were subtracted from the average z-score from H3K4me3 ChIP-chips.

RESULTS

The H3K4 methylation spreading observed in ctk1Δ cells is not caused by inappropriate RNA pol-II initiation in the body of genes. The level and pattern of cryptic initiation observed on the STE11 gene was no different between set2Δ and ctk1Δ cells (Figure 15). Yet, H3K4me3 levels were still significantly increased on the STE11 gene similar to results shown for FMP27. We detected no significant increase in cryptic initiation on the FMP27 gene in either set2Δ or ctk1Δ cells, despite significantly increased H3K4me3 levels (Figure 15).

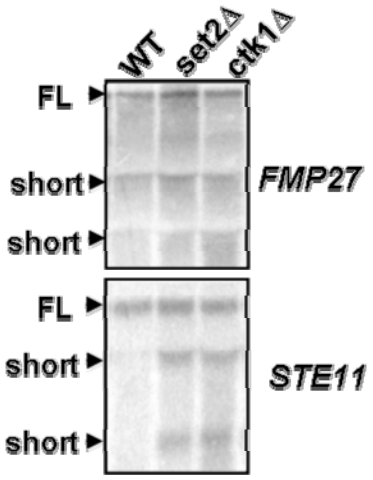


Figure 15 H3K4 methylation spreading observed in *ctk1Δ* cells cannot be explained by an increase in cryptic initiation or altered levels of transcription.

Total RNA purified from WT, *set2Δ* or *ctk1Δ* cells were used for Northern blot analysis and probed with radiolabeled PCR products corresponding to the gene. FL denotes full-length product and "short" denotes smaller RNA species that were observed. WT cells naturally exhibit some additional smaller mRNA products for the *FMP27* gene. Lanes were loaded with equivalent amounts of total RNA (assayed by ethidium staining, not shown). *STE11* was included as a positive control for cryptic initiation, which has been shown to occur within this gene upon deletion of *SET2* (Carrozza et al., 2005).

These data rule out large increases in RNA pol-II cryptic transcription initiation within the gene body as a cause for the spread of H3K4me3 in *ctk1*Δ strains.

The H3K4 methylation spreading observed in ctk1Δ cells occurs genome-wide and is restricted to the coding regions of RNA pol-II transcribed genes. To determine the extent that Ctk1p loss affects H3K4 methylation, I used a ChIP-chip approach to examine the genome-wide profiles of H3K4me3 in a wild-type and *CTK1* deletion strain. I used arrays that tiled continuously over the entire genome at a resolution of ~1 kb, and that additionally contained probes that covered all of chromosome III at a resolution of 200 bp (and at 100 bp resolution for 1/3 of the chromosome). I probed these arrays with DNA recovered from each of the three independent chromatin IPs that were performed by my collaborators. Upon analysis, I observed a significant change in the global patterns of H3K4me3 in *ctk1*Δ cells as compared to WT cells (Figure 16).

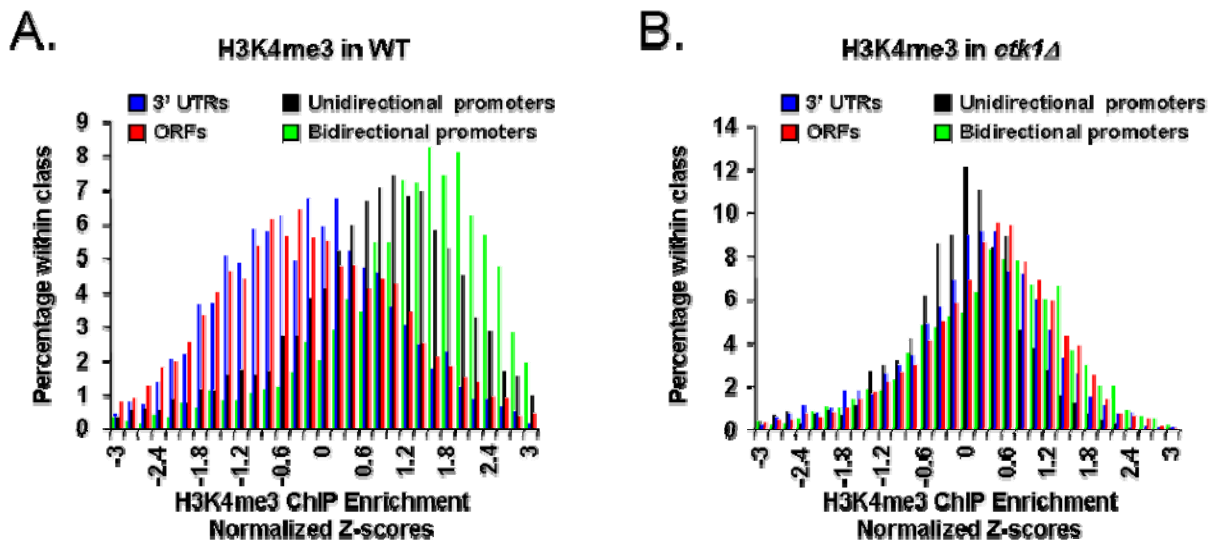


Figure 16 H3K4me3 spreading in *ctk1Δ* cells occurs genome-wide on the open reading frames of RNA pol-II regulated genes.

(A) The distribution of z-scores following H3K4me3 ChIPs from wild-type WCE. ORFs (red), 3' UTRs (blue), unidirectional promoters (black) and bidirectional promoters (green). Z-scores were normalized to histone H3 distribution as described (Materials and Methods). Promoters are clearly distinguished from ORFs and 3' UTRs (B) Same as in (A), but H3K4me3 ChIPs were performed with WCE from *ctk1Δ* cells.

In WT cells, and consistent with work by others (Liu et al., 2005; Ng et al., 2003b; Pokholok et al., 2005), I found that H3K4me3 was generally localized to the 5' end of loci transcribed by RNA pol-II (Figure 16A). However, in *ctk1*Δ cells, ORFs are enriched to a degree similar to that of promoters, consistent with a general spread of the mark into coding regions (Figure 16B). Note that the relative enrichment for any class is much less pronounced in the *ctk1*Δ ChIPs, which is also expected in the case of large-scale spreading.

I also asked directly if H3K4me3 signal spread into ORFs. The probes on our microarrays cover each ORF from start codon to stop codon. Since H3K4me3 in WT cells extends into the ORF by a fairly stereotypic distance regardless of ORF length, lower ratios will be reported for longer ORFs. This is because for increasingly long ORFs, an increasingly lower proportion of the ORF will harbor the H3K4me3 modification. This is precisely what is observed in the wild type strain (Figure 17A, black). However, in *ctk1*Δ cells (Figure 17A, red), this inverse correlation is not seen except for the longest ORFs, consistent with the spread of H3K4me3 well 3' of its normal boundary. I further examined the spreading using the high-resolution microarray coverage of chromosome III. In wild-type cells (and to a similar degree in *set2*Δ cells) the H3K4me3 signal drops sharply as a function of distance from transcriptional initiation, whereas in *ctk1*Δ mutants H3K4me3 levels remain uniform with increasing distance (Liu et al., 2005; Pokholok et al., 2005) from the site of initiation (Figure 17B).

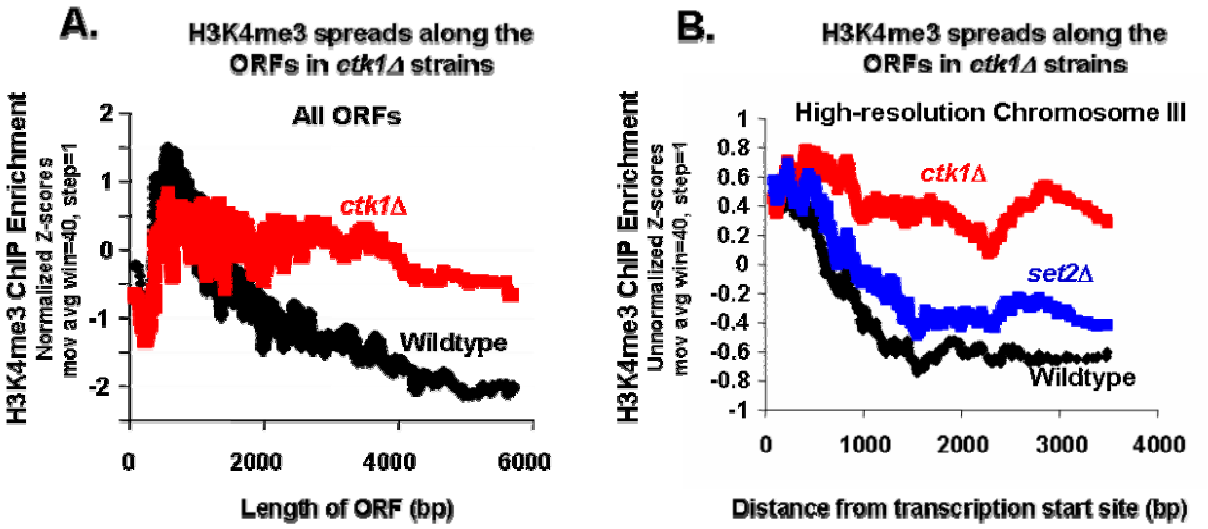


Figure 17 H3K4me3 spreads into the 3′ region of ORFs

(A) A moving average (window [win] = 40, step = 1) of normalized z-scores of H3K4me3 ChIP-chip data in wild type (black) and ctk1Δ (red), plotted as a function of ORF length. (B) A moving average (window [win] = 40, step = 1) of H3K4me3 ChIP-chip z-scores in wild type (black) set2Δ (blue) and ctk1Δ (red), plotted as a function of the distance from the transcription start site among genes greater than 1 kb in length and for which data was available (69 genes) on chromosome III. A high-resolution microarray covering chromosome III was used for this analysis (Materials and Methods) and the data plotted in this case were not normalized to histone occupancy.

These results confirm our findings on individual genes, show that the phenomenon occurs genome-wide, and strongly suggest that the spread of H3K4me3 occurs through inappropriate Set1p association with elongating RNA pol-II.

H3K4me3 was not observed on RNA pol-I and pol-III transcribed genes, or on other regions not transcribed by RNA pol-II including telomeres, HMR, and rDNA loci (Figure 18).This was confirmed by my collaborators using high-resolution analysis of a telomeric region on the right end of chromosome VI.

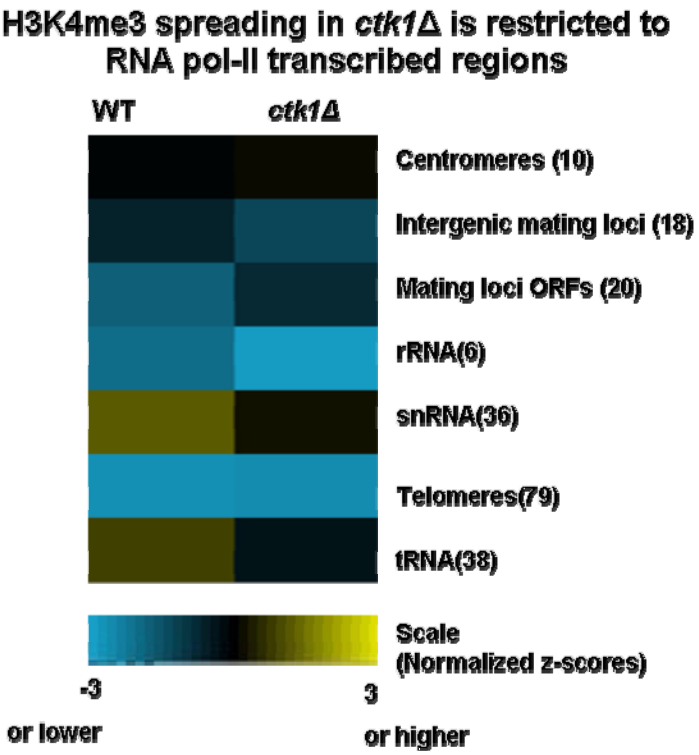


Figure 18 H3K4me3 spreading is restricted to RNA pol-II transcribed regions.

Color graphic representation of enrichment of H3K4me3 recorded from all arrayed elements in the indicated SGD functional class (labeled on right, number of arrayed elements analyzed in parenthesis). Colors (scale at bottom) represent the median of z-scores of all data points that were normalized to histone H3 distribution (see Material and Methods)

Chapter IV

SUMMARY

Since the discovery of covalent modifications on histones there has been tremendous advancement in our knowledge of the role of these modifications in gene expression and repression. Although the role of histone acetylation and its role in facilitating transcription was suggested a long time ago (Allfrey et al., 1964; Gorovsky et al., 1973), direct evidence of the role of acetylated nucleosomes in transcriptionally active genes was shown using antibodies that specifically recognize acetylated lysines (Hebbes et al., 1988). Further impetus in this field came by with the discovery of the first HAT as a homolog of yeast transcriptional coactivator Gcn5p (Brownell et al., 1996). Since then histone acetylation has been shown to be generally associated with activation. Histone acetylation is postulated to weaken histone-DNA (Hong et al., 1993; Steger and Workman, 1996) or nucleosome-nucleosome interactions (Fletcher and Hansen, 1996; Luger and Richmond, 1998). More recently, studies show that the SAGA and NuA4 HAT complexes strongly stimulated RSC's (a bromodomain containing ATPase) effect on elongation (Carey et al., 2006). Remodeling by RSC to facilitate passage of RNA pol-II is enhanced by the HATs.

A modification that has been the focus of many groups recently is histone methylation. While the enzymes responsible for these methylation events are being discovered and the properties of protein methylation have been employed to interpret the role of histone methylation, the function of the individual marks was not yet fully understood. With the development of antibodies that specifically recognize each of the modification states, it became possible to determine the *in vivo*

localization of these modifications and shed light on the role of these modifications in gene regulation.

Using chromatin immunoprecipitation to specifically isolate regions of the genome associated with a particular modification or protein of interest, several groups had shown histone acetylation to be predominantly enriched in the promoter region. Studies also showed H3K4me3 to be present in the 5′ end of coding regions and H3K36me was shown to be present at the 3' end of coding regions (Kizer et al., 2005; Krogan et al., 2003; Ng et al., 2003b; Pokholok et al., 2005; Santos-Rosa et al., 2002; Xiao et al., 2003).

In an effort to investigate the role of Set2p mediated H3K36me2, I set out to study the pattern of this modification on a global level. I used H3K36me2 specific antibody and whole genome yeast DNA microarrays to perform ChIP-chip experiments. I used DNA microarrays that cover the entire genome on a single slide for my studies. This was a significant improvement over many published studies that used arrays representing only the ORFs or only intergenic regions or that split ChIP samples and hybridized them independently to separate arrays representing only the ORFs or the intergenic regions. Use of a whole-genome array was essential to most of the conclusions in the study.

The next very important factor that I accounted for in the ChIP-chip studies were normalization to bulk nucleosome occupancy using data from H3 or H4-myc ChIPs that were performed in parallel from the same extract. This was important because studies had shown that nucleosome occupancy throughout the yeast genome is heterogeneous (Bernstein et al., 2004), and if left unaccounted for, misleading patterns could emerge. This was the first instance of modified-nucleosome ChIP data being normalized to apparent bulk nucleosome occupancy genome-wide. Nucleosome occupancy studies had revealed that nucleosomes are depleted from active

regulatory regions throughout the yeast genome and the level of nucleosome occupancy is inversely proportional to the frequency of transcription of the downstream gene.

My studies showed that H3K36me2 is present on all RNA pol-II transcribed regions. H3K36me2 is scarce or absent in upstream gene regulatory regions, telomeres, mating loci, and regions transcribed by RNA pol- III. This suggests that this mark is strictly associated with RNA pol-II during the elongation phase. The degree of H3K36me2 on the genes did not correlate with their frequency of transcription but a gene that was 'ON' tends to be methylated at H3K36 than a gene that is 'OFF'. My studies were the first to systematically reveal a chromatin mark that demarcates functionally distinct regions of the genome. The fact that the level of this mark did not correspond to the levels of transcription of genes *per se* suggested that it may not have a direct role in elongation. My studies also revealed that independent of the length of the ORF, H3K36me2 starts at a fixed distance from the transcription start site and goes on until the end of transcription. This suggests that there exists a clear point where Set2p's 'signal' is turned on and it gets functionally associated with RNA pol-II and starts methylating. Although Ctk1p mediated Ser-2 phosphorylation of CTD is a prerequisite for Set2's association with RNA pol-II what is not known is what triggers this switch from Ser-5 to Ser-2 phosphorylated form of CTD. This further emphasizes the importance of tight regulation of restricting this methyl mark only to the 3′ end of the genes.

The exact function of Set2p was not known until recently. Studies showed that the smaller subunit of the histone deacetylase complex, Rpd3(S) binds to H3K36me and deacetylases the nucleosomes behind the passing polymerase (Carrozza et al., 2005; Joshi and Struhl, 2005; Keogh et al., 2005). Deletion of Set2p resulted in inappropriate transcription initiation from within the genes (Carrozza et al., 2005; Joshi and Struhl, 2005). This phenotype is very similar to

that observed for histone chaperones like Spt6p and FACT complex subunit Spt16p. These factors are important transient disassembly and assembly of nucleosomes during transcription. In the absence of these factors, the nucleosome density is low on ORFs that are transcribing exposing the TATA boxes within the ORF. Thus while acetylation is a sign of transcription, subsequent deacetylation is required for maintaining transcription fidelity. H3K36me is important for ensuring transcription initiation at the 5′ end.

However, there are still some unresolved questions regarding the events transpiring during transcription. The life span of histone methylation is not known. Studies on H3K4me3 show that methylated nucleosomes are eventually diluted away by nucleosome replacement. It is not known if H3K36me is indeed a stable modification and is maintained through multiple rounds of transcription or if the methyl mark is removed in every round of transcription by the enzymes that are capable of reversing them (Klose et al 2006, Shi et al 2004, Tsukada et al 2006, Whetstine et al 2006).

My studies on the global distribution of H3K36ac showed that in contrast to H3K36me, H3K36 is acetylated in the regulatory regions of RNA pol-II transcribed genes. Although the role of this acetylation event is still unknown the fact that it is mediated by Gcn5p suggests it may have a role in transcription since Gcn5p is the HAT subunit of the SAGA complex that associates with RNA pol-II during transcription initiation. Whether there exists interplay between these two modifications remains to be seen. Studies on in mammals and fission yeast H have shown that H3K9 can be acetylated or methylated. However, acetylation has to be removed prior to methylation by *Su(var)3-9* in the promoters of genes (Schotta et al., 2002). An important question is whether acetylation of the residue at the promoters prevents methylation from spreading there. Since it is important for the cells to maintain the proper balance of acetylation

levels across the genome one mechanism of doing so is by restricting the methylation event to the 3′ end of genes. This would prevent hyperacetylation in the coding regions thus preventing aberrant initiation.

All these above studies suggest that since H3K36me is necessary for preventing aberrant transcription it can be considered repressive for transcription. It is possible that methylation therefore is a very crucial modification for an organism since it maintains a repressive and a stable state. The large numbers of transcription factors, chromatin remodelers are therefore required to transiently change this repressive state to an active one during transcription.

In contrast to H3K36me, H3K4me3 has been shown to be present at the 5′ end of actively transcribing genes. In an effort to understand the different mechanisms with which each of the modifications on H3K4 is regulated my collaborators performed a screen to test for factors regulating H3K4me1. They found that in *ctk1*Δ not only is there lower levels of H3K4me0 but there is an increased level of H3K4me3. I sought to investigate the extent of this spreading in the entire genome. Results showed that *ctk1*Δ caused genome wide spreading of H3K4me3 downstream into the 3′ end of ORFs. Since Set2p fails to associate with RNA pol-II in the absence of Ctk1p, we asked the question if the spreading of H3K4me3 was due to a loss of H3K36me. The results showed that this was not the case since in a *set2*Δ strain the pattern of H3K4me3 is similar to wildtype. This suggested that Ctk1p regulates this modification in a Set2p independent manner since only the loss H3K36me was not sufficient to cause the spreading of H3K4me3 in to the 3′ region of ORFs. I also investigated into the possibility whether this spreading effect was due to increased aberrant transcription in *ctk1*Δ. Northern blot results showed that this was not the case since the levels of aberrant transcripts were the same in *ctk1*Δ

and *set2*Δ strains. Although the mechanism is still unknown, these studies show that Ctk1p restricts H3K4me3 to the 5′ region at a global level.

In summary, my whole genome studies of a histone methylation mark, H3K36me2 was the first time that a histone modification pattern was analyzed accounting for the underlying nucleosome distribution. I showed that H3K36me2 was present in all the RNA pol-II transcribed regions including the 3′ end of ORFs and the 3′ UTRs. This study shows that H3K36me2 acts as a mark that defines chromatin context separating the genome into structurally and functionally distinct regions. The function of this modification was shown by other groups to prevent aberrant initiation within the ORFs by enabling deacetylation in this region thereby maintaining proper chromatin integrity during transcription.

Chapter V

PROJECTS IN PROGRESS

INTRODUCTION

Studies on global distribution of H3K36me2 and H3K36me3 had shown that this mark is present exclusively in the 3' end of coding regions thereby demarcating the genome functionally and structurally (my studies, Pokholok et al 2005), but its exact role in transcription was not known until some breakthrough studies that showed that methylated H3K36 is recognized by the chromodomain of Eaf3 in the Rpd3S histone deacetylase complex. Histone deacetylation by Rpd3S is an important event that is required for the suppression of aberrant internal transcription initiation (Carrozza et al., 2005; Joshi and Struhl, 2005; Keogh et al., 2005).

Since the presence of aberrant transcription initiation was shown on a few individual loci, it was not known how many genes in the entire yeast genome were influenced by the loss of H3K36me. Is there a particular class of genes that are affected, is there a correlation between the frequency of transcription of a gene and its ability to exhibit aberrant transcription? I sought to address these questions by directly comparing the expression profiles of *set2*Δ and wildtype strains.

We hypothesized that genes that are basally transcribed might be most likely affected by the loss of H3K36me. Genes transcribing rapidly have very low nucleosome density and therefore possess no barrier to binding of initiation complexes and are prone to internal initiation. We therefore hypothesized that yeast have developed a mechanism whereby genes that are actively transcribing prevent aberrant transcription either by having limiting amounts of initiation

complexes, by sequence specificity in the 5′ regions or by merely having high concentrations of RNA pol-II in the coding regions. Results showed that while steric hindrance by having high concentrations of RNA pol-II could be one of the mechanisms on some genes for preventing aberrant initiation, it is not the only mechanism.

MATERIALS AND METHODS

Northern blot analyses. Yeast Strains were grown at 30°C in YPR (1% yeast extract, 2% peptone, 2% raffinose) to an OD_{600} of 0.6-0.8. Galactose was added to the culture to a final concentration of 2% and samples were taken before addition of galactose and 0, 10, 20, 30, 60 and 90mins after addition of galactose. Total RNA was prepared by acid-phenol method as described (Xiao et al., 2003). For northern blotting, 30ug of each sample was subjected to electrophoresis in a formaldehyde-agarose gel, followed by transfer to a nylon membrane and crosslinked by UV irradiation. Hybridization was carried out in 6X SCP (2.5M NaCl, 0.7M Na_2HPO_4, 0.02M EDTA), 1% Sarkosyl and 0.1 mg/ml of salmon sperm DNA. Probes (full-length *STE11* and *FLO8*) were generated by PCR.

Expression Data. WT (BY4741) and *set2*Δ (BY4741) strains were grown at 30°C in YPD (1% yeast extract, 2% peptone, 2% dextrose) to an OD_{600} of 0.6-0.8. Total RNA was prepared by acid-phenol method as described (Xiao et al 2003). Ds cDNA was prepared using Invitrogen SuperScript™ Double-Stranded cDNA synthesis kit (Cat No. 11917-010). The DNA was

labeled, hybridized and scanned by Nimblegen, Inc. The arrays used for the analysis were high resolution CGH arrays.

RESULTS

Recent studies have shown that absence of Set2p-mediated H3K36me, and loss of function of Rpd3 complex on individual genes, causes an increase in histone acetylation in the coding regions of genes (Carrozza et al., 2005; Joshi and Struhl, 2005; Keogh et al., 2005). Some of these genes exhibited inappropriate transcription initiating from within the gene body. I wanted to ask the question how many genes in the entire yeast genome are affected by deletion of Set2p by directly comparing the transcript levels in *set2*Δ and WT strains. We hypothesized that due to the presence of new shorter transcripts, in addition to the full length transcripts in the *set2*Δ strain, we would expect to see increased ratios on the 3′ end of the ORFs.

I extracted total RNA from three independent cultures each of *set2*Δ and WT strains and generated double stranded cDNA (see materials and methods). The cDNA was sent to Nimblegen systems, Inc. for hybridization on high resolution yeast whole genome arrays. Three independent biological replicates were performed one of which was labeled as a dye swap. Upon preliminary analysis, I found that when compared to WT, both the positive control genes, *STE11* and *FLO8* exhibited a clear increase in signal in the *set2*Δ strain at the 3′ end (data not shown). My plans are to further analyze this data by taking the average of the three replicates and looking for genes that show this difference.

We hypothesized that infrequently transcribed genes have a higher tendency to exhibit aberrant transcription initiation than a gene that is transcribing a high rate. Genes transcribing at low levels have low density of RNA pol-II and higher levels of nucleosomes that need to be

deacetylated to prevent aberrant transcription in comparison to highly transcribing genes like the heat shock genes that do not have restrictive chromatin owing to excessive loss of nucleosomes. We hypothesize that yeast have developed a mechanism whereby genes that are actively transcribing prevent aberrant transcription either by having limiting amounts of initiation complexes, by sequence specificity in the 5′ regions or by merely having high concentrations of RNA pol-II in the coding regions. I set out to test this hypothesis by artificially inducing a gene that has been shown to exhibit aberrant initiation and investigate if cells can evade the requirement of Set2p.

The endogenous promoter of two candidate genes, *STE11* and *FLO8* were replaced with the *GAL1* promoter in *spt6*ts and WT background. Spt6p is a transcription factor that is involved in nucleosome disassembly and assembly during transcription. *spt6-1004*, the temperature sensitive strain used in these studies, has been shown to exhibit aberrant transcription at permissive temperature. I therefore used this strain as a positive control.

Cells were grown in raffinose up to an OD of 0.8-1.0. Galactose was then added to a final concentration of 2% and grown at 30^{o}C for 90min. Samples were taken before and after addition of galactose for RNA extraction. As shown in the Figure 19, *spt6*ts (*GAL1-STE11*) strain exhibits two short transcripts before the addition of galactose. These shorter bands are lost 90min after the addition of galactose thus supporting our hypothesis.

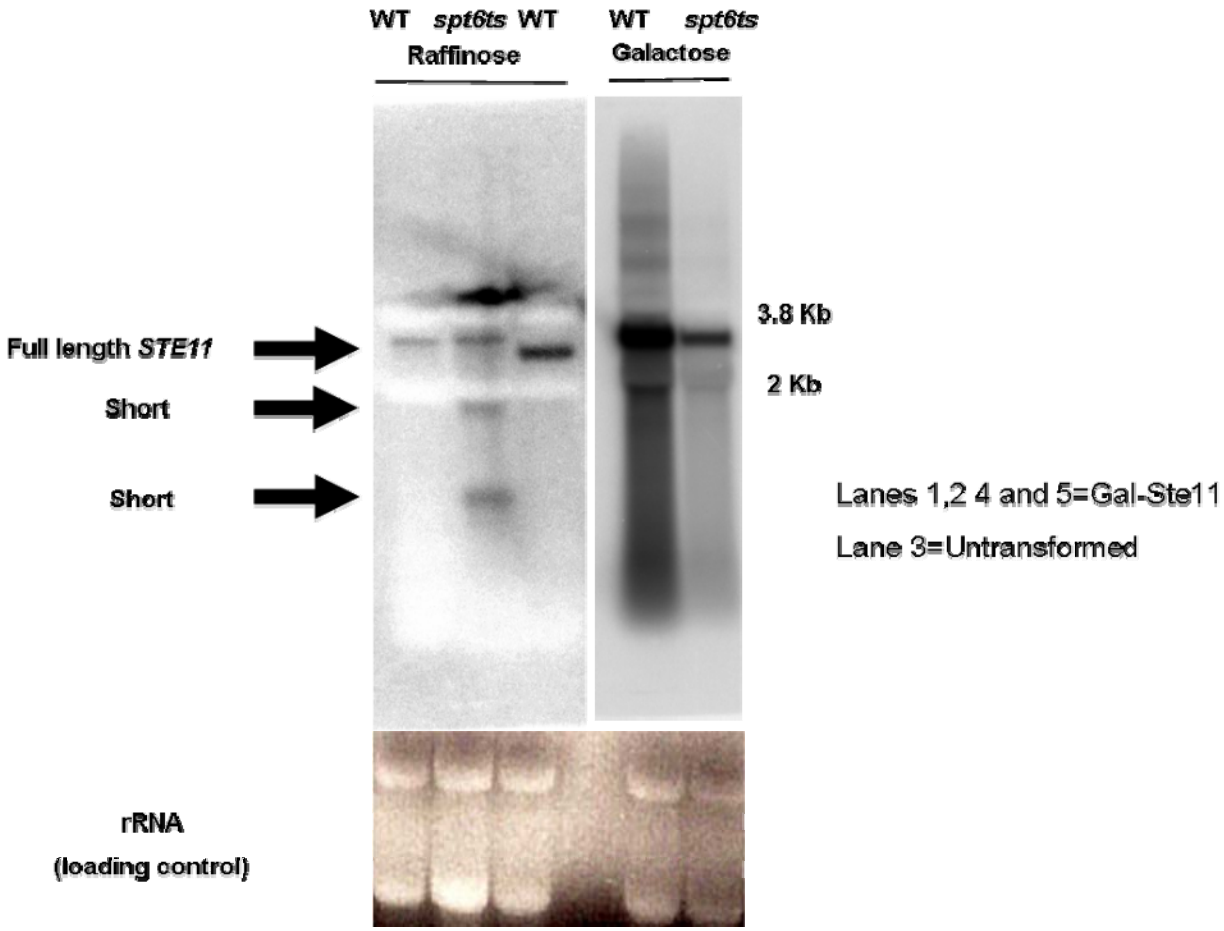


Figure 19 Higher levels of transcription result in loss of aberrant initiation in the absence of Spt6p.

In the presence of raffinose, the spt6ts strain exhibits the short aberrant transcripts in addition to the full length transcript. Upon induction with galactose, the shorter transcripts are lost indicating loss of aberrant initiation. The slightly longer size of the full length transcript in the transformed strains (lanes 1, 2, 4 and 5) is due to the fact that the presence of *GAL1* promoter results in transcription to initiate upstream of the regular site.

I next performed a time course experiment to determine at what point after the addition of galactose the shorter transcripts are lost. As shown in the Figure 20, aberrant transcription in *STE11* is lost around 30mins after its induction. I performed the same experiment using WT (*GAL-FLO8*) and *spt6ts* (*GAL-FLO8*). To my surprise, around 10mins after induction, the WT (*GAL-FLO8*) started exhibiting multiple shorter transcripts (Figure 21). The results from the

spt6ts (*GAL-FLO8*) were inconclusive and I propose to repeat this experiment in order to examine the pattern of aberrant initiation in this strain.

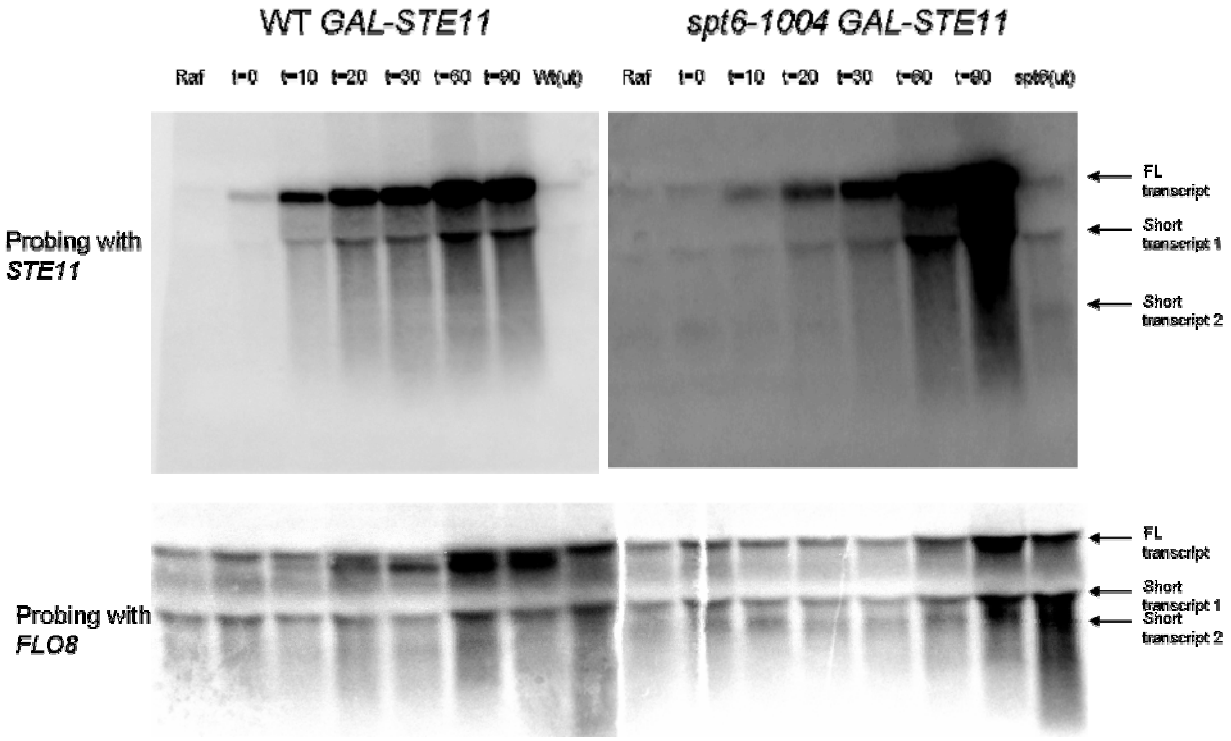


Figure 20 Time course of galactose induction of Ste11 shows aberrant transcription is lost around 30mins after induction.

Probing with *STE11* shows that in the *spt6ts* though the levels of full length transcript increases due to induction, the shorter transcript 2 is lost around 30mins after induction. The membranes were stripped and reprobed with *FLO8* as a control for RNA levels. The darker bands in the t=60 and t=90 lanes are due to incomplete stripping of the *STE11* transcripts.

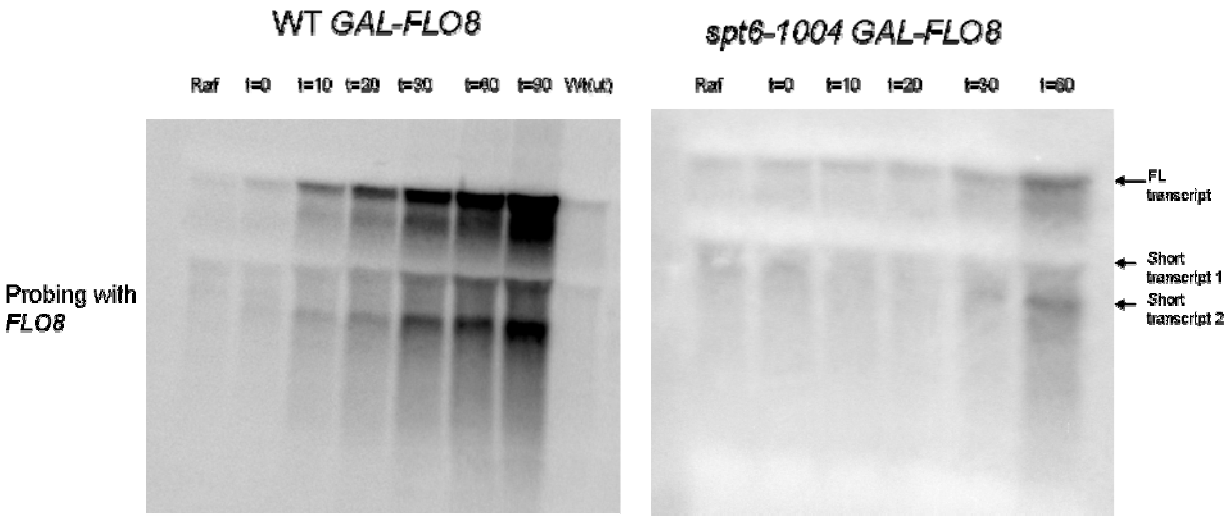


Figure 21 Time course of galactose induction of *FLO8* shows aberrant transcription is induced in a wildtype background.

Probing with *FLO8* shows that in the wildtype background, shorter transcripts appear about 10mins after induction. The results from the *spt6ts* strain background were inconclusive.

DISCUSSION

It was recently shown that histone deacetylase complex Rpd3(S) binds to H3K36me2 in the wake of the passing polymerase and maintains proper acetylation levels in the coding regions. Deletion of *SET2* was shown to result in shorter transcripts initiating from the 3′ region of genes. Set2p may therefore play a role in maintaining proper chromatin structure after RNA pol-II has passed by restricting transcription initiation to the 5′ region. However, we hypothesized that yeast have developed a mechanism whereby genes that are actively transcribing prevent aberrant transcription either by having limiting amounts of initiation complexes, by sequence specificity in the 5′ regions or by merely having high concentrations of RNA pol-II in the coding regions.

I tested out the hypothesis that one of the mechanisms by which aberrant transcription in actively transcribing genes is prevented is by having high concentrations of RNA pol-II. The RNA pol-II molecules thus cause a steric hindrance and prevent initiation complexes from accumulating in the internal start site. In order to test this hypothesis on two candidate genes, I created inducible strains *GAL-STE11* and *GAL-FLO8* in a wildtype background and in a background wherein the original copy of the histone chaperone/elongation factor, Spt6p had been replaced with a mutant form. Both the candidate genes exhibited aberrant transcription in this mutant *spt6* background.

The results showed that in the mutant *spt6* background, *STE11* exhibits inappropriate transcription that is lost around 30mins after its induction by galactose. This coincides with the time around when the concentration of RNA pol-II become saturated on *GAL1* gene upon induction with galactose (Xiao et al., 2005). Considering their similarity in size it can be assumed that the levels of RNA pol-II come to a saturating level on *STE11* around the same time period.

However, contrary to *STE11*, the *FLO8* gene started exhibiting aberrant transcripts in the wildtype background upon induction. This result led me to conclude that the aberrant transcription in these genes is perhaps regulated by different mechanisms. I hope to address this issue using the expression data. There could be several possibilities that can be hypothesized. *FLO8* has an internal TATA box that has been characterized to initiate transcription (Carrozza et al., 2005; Kaplan et al., 2003) whereas *STE11* has not been shown to have a functional internal TATA element. It is also possible that the fundamental nucleosome organization is different in *STE11* and *FLO8.* It will be interesting to see what factors control regulation of aberrant transcription initiation. The function of the aberrant transcripts is still not known although studies show that they might be polyadenylated (Carrozza et al., 2005). It is possible that these

transcripts are translated to proteins and may indeed have a functional role. Multiple splice forms of proteins are known to widely exist in mammals and other organisms. The hypothesis that Set2p regulates generation of aberrant transcripts in basally transcribed genes is substantial since majority of genes in yeast and higher eukaryotes are transcribed at basal levels and therefore organisms have developed multiple layers of mechanisms to ensure transcription fidelity.

REFERENCES

Adkins, M. W., and Tyler, J. K. (2004). The histone chaperone Asf1p mediates global chromatin disassembly in vivo. J Biol Chem *279*, 52069-52074.

Allfrey, V. G., Faulkner, R., and Mirsky, A. E. (1964). Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. Proc Natl Acad Sci U S A *51*, 786-794.

Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial distribution of di-and tri-methyl lysine 36 of histone H3 at active genes. J Biol Chem.

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Belotserkovskaya, R., and Reinberg, D. (2004). Facts about FACT and transcript elongation through chromatin. Curr Opin Genet Dev *14*, 139-146.

Bergman, L. W., and Kramer, R. A. (1983). Modulation of chromatin structure associated with derepression of the acid phosphatase gene of Saccharomyces cerevisiae. J Biol Chem *258*, 7223-7227.

Bernstein, B. E., Humphrey, E. L., Erlich, R. L., Schneider, R., Bouman, P., Liu, J. S., Kouzarides, T., and Schreiber, S. L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci U S A *99*, 8695-8700.

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

Boa, S., Coert, C., and Patterton, H. G. (2003). Saccharomyces cerevisiae Set1p is a methyltransferase specific for lysine 4 of histone H3 and is required for efficient gene expression. Yeast *20*, 827-835.

Boeger, H., Griesenbeck, J., Strattan, J. S., and Kornberg, R. D. (2003). Nucleosomes unfold completely at a transcriptionally active promoter. Mol Cell *11*, 1587-1598.

Boeger, H., Griesenbeck, J., Strattan, J. S., and Kornberg, R. D. (2004). Removal of promoter nucleosomes by disassembly rather than sliding in vivo. Mol Cell *14*, 667-673.

Bohlander, S. K., Espinosa, R., 3rd, Le Beau, M. M., Rowley, J. D., and Diaz, M. O. (1992). A method for the rapid sequence-independent amplification of microdissected chromosomal material. Genomics *13*, 1322-1324.

Briggs, S. D., Bryk, M., Strahl, B. D., Cheung, W. L., Davie, J. K., Dent, S. Y., Winston, F., and Allis, C. D. (2001). Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. Genes Dev *15*, 3286-3295.

Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., and Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843-851.

Bryk, M., Briggs, S. D., Strahl, B. D., Curcio, M. J., Allis, C. D., and Winston, F. (2002). Evidence that Set1, a factor required for methylation of histone H3, regulates rDNA silencing in S. cerevisiae by a Sir2-independent mechanism. Curr Biol *12*, 165-170.

Buck, M. J., and Lieb, J. D. (2004). ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics *83*, 349-360.

Carey, M., Li, B., and Workman, J. L. (2006). RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. Mol Cell *24*, 481-487.

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W. J., Anderson, S., Yates, J., Washburn, M. P., and Workman, J. L. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Causton, H. C., Ren, B., Koh, S. S., Harbison, C. T., Kanin, E., Jennings, E. G., Lee, T. I., True, H. L., Lander, E. S., and Young, R. A. (2001). Remodeling of yeast genome expression in response to environmental changes. Mol Biol Cell *12*, 323-337.

Cheung, P., Tanner, K. G., Cheung, W. L., Sassone-Corsi, P., Denu, J. M., and Allis, C. D. (2000). Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. Mol Cell *5*, 905-915.

Cho, E. J., Kobor, M. S., Kim, M., Greenblatt, J., and Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes Dev *15*, 3319-3329.

Clements, A., Poux, A. N., Lo, W. S., Pillus, L., Berger, S. L., and Marmorstein, R. (2003). Structural basis for histone and phosphohistone binding by the GCN5 histone acetyltransferase. Mol Cell *12*, 461-473.

Deckert, J., and Struhl, K. (2002). Targeted recruitment of Rpd3 histone deacetylase represses transcription by inhibiting recruitment of Swi/Snf, SAGA, and TATA binding protein. Mol Cell Biol *22*, 6458-6470.

Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M., Garcia, B. A., Shabanowitz, J., Hunt, D. F., Funabiki, H., and Allis, C. D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature *438*, 1116-1122.

Fletcher, T. M., and Hansen, J. C. (1996). The nucleosomal array: structure/function relationships. Crit Rev Eukaryot Gene Expr *6*, 149-188.

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D., and Brown, P. O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell *11*, 4241-4257.

Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M*., et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.

Gerton, J. L., DeRisi, J., Shroff, R., Lichten, M., Brown, P. O., and Petes, T. D. (2000). Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *97*, 11383-11390.

Gorovsky, M. A., Pleger, G. L., Keevert, J. B., and Johmann, C. A. (1973). Studies on histone fraction F2A1 in macro- and micronuclei of Tetrahymena pyriformis. J Cell Biol *57*, 773-781.

Govind, C. K., Zhang, F., Qiu, H., Hofmeyer, K., and Hinnebusch, A. G. (2007). Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions. Mol Cell *25*, 31-42.

Grant, P. A., Eberharter, A., John, S., Cook, R. G., Turner, B. M., and Workman, J. L. (1999). Expanded lysine acetylation specificity of Gcn5 in native complexes. J Biol Chem *274*, 5895-5900.

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.

Hanlon, S. E., and Lieb, J. D. (2004). Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. Curr Opin Genet Dev *14*, 697-705.

Hebbes, T. R., Thorne, A. W., and Crane-Robinson, C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. Embo J *7*, 1395-1402.

Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Green, M. R., Golub, T. R., Lander, E. S., and Young, R. A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell *95*, 717-728.

Hong, L., Schroth, G. P., Matthews, H. R., Yau, P., and Bradbury, E. M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. J Biol Chem *268*, 305-314.

Hurowitz, E. H., and Brown, P. O. (2003). Genome-wide analysis of mRNA lengths in Saccharomyces cerevisiae. Genome Biol *5*, R2.

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

Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M., and Brown, P. O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature *409*, 533-538.

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jenuwein, T., Laible, G., Dorn, R., and Reuter, G. (1998). SET domain proteins modulate chromatin domains in eu- and heterochromatin. Cell Mol Life Sci *54*, 80-93.

Jones, J. C., Phatnani, H. P., Haystead, T. A., MacDonald, J. A., Alam, S. M., and Greenleaf, A. L. (2004). C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats. J Biol Chem *279*, 24957-24964.

Joshi, A. A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Mol Cell *20*, 971-978.

Kaplan, C. D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science *301*, 1096-1099.

Keogh, M. C., Kurdistani, S. K., Morris, S. A., Ahn, S. H., Podolny, V., Collins, S. R., Schuldiner, M., Chin, K., Punna, T., Thompson, N. J.*, et al.* (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell *123*, 593-605.

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.

Kizer, K. O., Phatnani, H. P., Shibata, Y., Hall, H., Greenleaf, A. L., and Strahl, B. D. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Mol Cell Biol *25*, 3305-3316.

Knezetic, J. A., and Luse, D. S. (1986). The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. Cell *45*, 95-104.

Komarnitsky, P., Cho, E. J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev *14*, 2452-2460.

Kornberg, R. D., and Thomas, J. O. (1974). Chromatin structure; oligomers of the histones. Science *184*, 865-868.

Kristjuhan, A., and Svejstrup, J. Q. (2004). Evidence for distinct mechanisms facilitating transcript elongation through chromatin in vivo. Embo J *23*, 4243-4252.

Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., and Greenblatt, J. F. (2002). RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. Mol Cell Biol *22*, 6979-6992.

Krogan, N. J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D. P., Beattie, B. K., Emili, A., Boone, C.*, et al.* (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol Cell Biol *23*, 4207-4218.

Kuo, M. H., and Allis, C. D. (1999). In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. Methods *19*, 425-433.

Kuo, M. H., Brownell, J. E., Sobel, R. E., Ranalli, T. A., Cook, R. G., Edmondson, D. G., Roth, S. Y., and Allis, C. D. (1996). Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. Nature *383*, 269-272.

Kuo, M. H., vom Baur, E., Struhl, K., and Allis, C. D. (2000). Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription. Mol Cell *6*, 1309-1320.

Kwon, J., Morshead, K. B., Guyon, J. R., Kingston, R. E., and Oettinger, M. A. (2000). Histone acetylation and hSWI/SNF remodeling act in concert to stimulate V(D)J cleavage of nucleosomal DNA. Mol Cell *6*, 1037-1048.

Lacoste, N., Utley, R. T., Hunter, J. M., Poirier, G. G., and Cote, J. (2002). Disruptor of telomeric silencing-1 is a chromatin-specific histone H3 methyltransferase. J Biol Chem *277*, 30421-30424.

Landry, J., Sutton, A., Hesman, T., Min, J., Xu, R. M., Johnston, M., and Sternglanz, R. (2003). Set2-catalyzed methylation of histone H3 represses basal expression of GAL4 in Saccharomyces cerevisiae. Mol Cell Biol *23*, 5972-5978.

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

Lee, D. Y., Hayes, J. J., Pruss, D., and Wolffe, A. P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell *72*, 73-84.

Lee, J. M., and Greenleaf, A. L. (1991). CTD kinase large subunit is encoded by CTK1, a gene required for normal growth of Saccharomyces cerevisiae. Gene Expr *1*, 149-167.

Levy, A., and Noll, M. (1981). Chromatin fine structure of active and repressed genes. Nature *289*, 198-203.

Li, B., Howe, L., Anderson, S., Yates, J. R., 3rd, and Workman, J. L. (2003). The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem *278*, 8897-8903.

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

Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D., and Patel, D. J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature *442*, 91-95.

Li, J., Moazed, D., and Gygi, S. P. (2002). Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. J Biol Chem *277*, 49383-49388.

Licatalosi, D. D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J. B., and Bentley, D. L. (2002). Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. Mol Cell *9*, 1101-1111.

Lieb, J. D., Liu, X., Botstein, D., and Brown, P. O. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet *28*, 327-334.

Liu, C. L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S. L., Friedman, N., and Rando, O. J. (2005). Single-nucleosome mapping of histone modifications in S. cerevisiae. PLoS Biol *3*, e328.

Lo, W. S., Trievel, R. C., Rojas, J. R., Duggan, L., Hsu, J. Y., Allis, C. D., Marmorstein, R., and Berger, S. L. (2000). Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. Mol Cell *5*, 917-926.

Loden, M., and van Steensel, B. (2005). Whole-genome views of chromatin structure. Chromosome Res *13*, 289-298.

Lorch, Y., LaPointe, J. W., and Kornberg, R. D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell *49*, 203-210.

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.

Luger, K., and Richmond, T. J. (1998). The histone tails of the nucleosome. Curr Opin Genet Dev *8*, 140-146.

Mai, X., Chou, S., and Struhl, K. (2000). Preferential accessibility of the yeast his3 promoter is determined by a general property of the DNA sequence, not by specific elements. Mol Cell Biol *20*, 6668-6676.

Martens, J. A., and Winston, F. (2003). Recent advances in understanding chromatin remodeling by Swi/Snf complexes. Curr Opin Genet Dev *13*, 136-142.

Mason, P. B., and Struhl, K. (2003). The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. Mol Cell Biol *23*, 8323-8333.

Masumoto, H., Hawke, D., Kobayashi, R., and Verreault, A. (2005). A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. Nature *436*, 294-298.

Morse, R. H. (2000). RAP, RAP, open up! New wrinkles for RAP1 in yeast. Trends Genet *16*, 51-53.

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

Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., and Grewal, S. I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110-113.

Nelson, C. J., Santos-Rosa, H., and Kouzarides, T. (2006). Proline isomerization of histone H3 regulates lysine methylation and gene expression. Cell *126*, 905-916.

Ng, H. H., Ciccone, D. N., Morshead, K. B., Oettinger, M. A., and Struhl, K. (2003a). Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: a potential mechanism for position-effect variegation. Proc Natl Acad Sci U S A *100*, 1820-1825.

Ng, H. H., Feng, Q., Wang, H., Erdjument-Bromage, H., Tempst, P., Zhang, Y., and Struhl, K. (2002). Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. Genes Dev *16*, 1518-1527.

Ng, H. H., Robert, F., Young, R. A., and Struhl, K. (2003b). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell *11*, 709-719.

Nightingale, K. P., Wellinger, R. E., Sogo, J. M., and Becker, P. B. (1998). Histone acetylation facilitates RNA polymerase II transcription of the Drosophila hsp26 gene in chromatin. Embo J *17*, 2865-2876.

Nourani, A., Utley, R. T., Allard, S., and Cote, J. (2004). Recruitment of the NuA4 complex poises the PHO5 promoter for chromatin remodeling and activation. Embo J *23*, 2597-2607.

Orphanides, G., Wu, W. H., Lane, W. S., Hampsey, M., and Reinberg, D. (1999). The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. Nature *400*, 284-288.

Pena, P. V., Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V., Gozani, O., Zhao, R., and Kutateladze, T. G. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature *442*, 100-103.

Placek, B. J., Harrison, L. N., Villers, B. M., and Gloss, L. M. (2005). The H2A.Z/H2B dimer is unstable compared to the dimer containing the major H2A isoform. Protein Sci *14*, 514-522.

Pokholok, D. K., Hannett, N. M., and Young, R. A. (2002). Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. Mol Cell *9*, 799-809.

Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E.*, et al.* (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517-527.

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.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Reid, J. L., Moqtaderi, Z., and Struhl, K. (2004). Eaf3 regulates the global pattern of histone acetylation in Saccharomyces cerevisiae. Mol Cell Biol *24*, 757-764.

Reinke, H., and Horz, W. (2003). Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. Mol Cell *11*, 1599-1607.

Robert, F., Pokholok, D. K., Hannett, N. M., Rinaldi, N. J., Chandy, M., Rolfe, A., Workman, J. L., Gifford, D. K., and Young, R. A. (2004). Global position and recruitment of HATs and HDACs in the yeast genome. Mol Cell *16*, 199-209.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W. W., Wilm, M., Aasland, R., and Stewart, A. F. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. Embo J *20*, 7137-7148.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Schaft, D., Roguev, A., Kotovic, K. M., Shevchenko, A., Sarov, M., Neugebauer, K. M., and Stewart, A. F. (2003). The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation. Nucleic Acids Res *31*, 2475-2482.

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., and Reuter, G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. Embo J *21*, 1121-1131.

Schwabish, M. A., and Struhl, K. (2004). Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. Mol Cell Biol *24*, 10111-10117.

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

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

Sekinger, E. A., Moqtaderi, Z., and Struhl, K. (2005b). Intrinsic Histone-DNA Interactions and Low Nucleosome Density Are Important for Preferential Accessibility of Promoter Regions in Yeast. Molecular Cell *18*, In Press.

Sewack, G. F., Ellis, T. W., and Hansen, U. (2001). Binding of TATA binding protein to a naturally positioned nucleosome is facilitated by histone acetylation. Mol Cell Biol *21*, 1404-1415.

Shi, X., Hong, T., Walter, K. L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M. R.*, et al.* (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature *442*, 96-99.

Squazzo, S. L., Costa, P. J., Lindstrom, D. L., Kumer, K. E., Simic, R., Jennings, J. L., Link, A. J., Arndt, K. M., and Hartzog, G. A. (2002). The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. Embo J *21*, 1764-1774.

Stapleton, A., and Petes, T. D. (1991). The Tn3 beta-lactamase gene acts as a hotspot for meiotic recombination in yeast. Genetics *127*, 39-51.

Steger, D. J., Eberharter, A., John, S., Grant, P. A., and Workman, J. L. (1998). Purified histone acetyltransferase complexes stimulate HIV-1 transcription from preassembled nucleosomal arrays. Proc Natl Acad Sci U S A *95*, 12924-12929.

Steger, D. J., and Workman, J. L. (1996). Remodeling chromatin structures for transcription: what happens to the histones? Bioessays *18*, 875-884.

Sterner, D. E., Lee, J. M., Hardin, S. E., and Greenleaf, A. L. (1995). The yeast carboxyl-terminal repeat domain kinase CTDK-I is a divergent cyclin-cyclin-dependent kinase complex. Mol Cell Biol *15*, 5716-5724.

Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Strahl, B. D., Grant, P. A., Briggs, S. D., Sun, Z. W., Bone, J. R., Caldwell, J. A., Mollah, S., Cook, R. G., Shabanowitz, J., Hunt, D. F., and Allis, C. D. (2002). Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol Cell Biol *22*, 1298-1306.

Strahl, B. D., Ohba, R., Cook, R. G., and Allis, C. D. (1999). Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. Proc Natl Acad Sci U S A *96*, 14967-14972.

Studitsky, V. M., Walter, W., Kireeva, M., Kashlev, M., and Felsenfeld, G. (2004). Chromatin remodeling by RNA polymerases. Trends Biochem Sci *29*, 127-135.

Svaren, J., and Horz, W. (1997). Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. Trends Biochem Sci *22*, 93-97.

Turner, B. M. (2000). Histone acetylation and an epigenetic code. Bioessays *22*, 836-845.

van holde, K. E. (1988). Chromatin, Springer, New York).

van Steensel, B., and Henikoff, S. (2003). Epigenomic profiling using microarrays. Biotechniques *35*, 346-350, 352-344, 356-347.

Vettese-Dadey, M., Grant, P. A., Hebbes, T. R., Crane- Robinson, C., Allis, C. D., and Workman, J. L. (1996). Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. Embo J *15*, 2508-2518.

Wang, L., Mizzen, C., Ying, C., Candau, R., Barlev, N., Brownell, J., Allis, C. D., and Berger, S. L. (1997). Histone acetyltransferase activity is conserved between yeast and human GCN5 and is required for complementation of growth and transcriptional activation. Mol Cell Biol *17*, 519-527.

Workman, J. L., and Roeder, R. G. (1987). Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. Cell *51*, 613-622.

Wu, T. C., and Lichten, M. (1994). Meiosis-induced double-strand break sites determined by yeast chromatin structure. Science *263*, 515-518.

Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M., Tackett, A. J., Chait, B. T., Badenhorst, P.*, et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86-90.

Xiao, T., Hall, H., Kizer, K. O., Shibata, Y., Hall, M. C., Borchers, C. H., and Strahl, B. D. (2003). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev *17*, 654-663.

Xiao, T., Kao, C. F., Krogan, N. J., Sun, Z. W., Greenblatt, J. F., Osley, M. A., and Strahl, B. D. (2005). Histone H2B ubiquitylation is associated with elongating RNA polymerase II. Mol Cell Biol *25*, 637-651.

Ye, J., Ai, X., Eugeni, E. E., Zhang, L., Carpenter, L. R., Jelinek, M. A., Freitas, M. A., and Parthun, M. R. (2005). Histone H4 lysine 91 acetylation a core domain modification associated with chromatin assembly. Mol Cell *18*, 123-130.

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

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

Zegerman, P., Canas, B., Pappin, D., and Kouzarides, T. (2002). Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. J Biol Chem *277*, 11621-11624.

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

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev *15*, 2343-2360.