# Epigenetic regulation of gene transcription: Jhdm1b characterization and genome-wide localization of H2A ubiquitylation

**By**
**Eric M. Kallin**

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 of Genetics and Molecular Biology.

Chapel Hill
2009

Approved by:

Dr. Yi Zhang

Dr. Jenny Ting

Dr. Beverly Koller

Dr. Jason Lieb

Dr. Bernard Weissman

## ABSTRACT

Eric M. Kallin

Epigenetic regulation of gene transcription: Jhdm1b characterization and genome-wide localization of H2A ubiquitylation

(Under the direction of Dr. Yi Zhang)

Histone methylation and ubiquitylation play important roles in regulating gene expression and form part of the epigenetic memory system that regulates cellular fate. Two general classes of enzymes linked to covalent modification are the JmjC domain-containing proteins capable of histone demethylation and the E3 ubiquitin ligases. Here we report the evolutionary conservation of JmjC domain-containing proteins with the goal of predicting active demethylases. We demonstrate that one of these proteins, Jhdm1b, is an H3K36 demethylase both *in vitro* and *in vivo*. Knock-down of Jhdm1b in primary MEFs inhibits cell proliferation and induces cellular senescence in a pRb and p53 pathway-dependent manner that is mediated through direct de-repression of $p15^{Ink4b}$. Jhdm1b binds to the $p15^{Ink4b}$ promoter and maintains low levels of local H3K36me2. Importantly, knock-down followed by rescue experiments demonstrate that this function is mediated through an active JmjC domain.

Transcriptional regulation by PcG proteins is achieved at least partly through the PRC1-mediated ubiquitylation of histone H2A (uH2A). Bmi1 has been demonstrated to be critical for H2A ubiquitylation. Although recent studies have revealed the genome-wide binding patterns of some of the PRC1 components there have been no reports describing

genome-wide localization of uH2A. Using the recently developed ChIP-Seq technology, here we report genome wide localization of the Bmi1-dependent uH2A mark in MEF cells. Gene promoter averaging analysis indicates a peak of uH2A just inside the transcription start site (TSS) of well annotated genes that is enriched at promoters containing H3K27me3. Peak finding reveals regions of local uH2A enrichment throughout the mouse genome, including almost 700 gene promoters. Genes with promoter peaks of uH2A exhibit lower level expression when compared to genes that do not contain promoter peaks of uH2A. We demonstrate that genes with uH2A peaks have increased expression upon Bmi1 knockout and local enrichment of uH2A is not limited to regions containing the H3K27me3 mark. Our work not only reveals Bmi1-dependent H2A ubiquitylation but also suggests that uH2A targeting in differentiated cells may employ a different mechanism from that in ES cells.

To my wonderfully supportive and devotedly loving family, I could not have done this without you. Nor could I ever express enough gratitude for the role you have played in shaping who I am today.

## ACKNOWLEDGEMENTS

I would be remiss to not acknowledge the significant role of following persons in shaping this work: Thank you to Dr. Yi Zhang for not only his guidance and support but also his critical eye that taught me the importance of being attentive and prepared. Thank you to the members of my committee, Dr. Buddy Weissman, Dr. Jason Lieb, Dr. Jenny Ting and Dr. Bev Koller for their generous investment of time and expertise. I look forward to the day that I might achieve your levels of sharpness and critical reasoning. Thank you to past and present members of the Zhang Lab for sharing their support, technical expertise, and friendship over these several years. Thank you to Susan Wu for being a confidant and a strong shoulder, but most of all, for pretending to laugh at my jokes. Finally, thank you to three of the most wonderful people I have ever met, a group I am lucky to call my closest friends. Sarah, Lorene, and Brett, you will always be held in a special part of my heart.

## TABLE OF CONTENTS

## LIST OF TABLES

Page

## LIST OF FIGURES

Page

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AEBP2 | Adipocyte enhancer binding protein 2 |
| AR | Androgen receptor |
| BMI1 | B lymphoma Mo-MLV insertion region 1 |
| ChIP | Chromatin immunoprecipitation |
| ChIP-chip | ChIP coupled to genomic microarrays |
| ChIP-seq | ChIP coupled to high throughput sequencing |
| EED | Embryonic ectoderm development |
| EZH2 | Enhancer of zeste homolog 2 |
| E(Z) | Enhancer of zeste |
| FIH | Factor inhibiting hypoxia |
| HAT | Histone acetyltransferase |
| HDAC | Histone deacetylase |
| HMT | Histone methyltransferase |
| HOX | Homeobox |
| HPH2 | Polyhomeotic homolog 2 |
| IP | Immunoprecipitation |
| JHDM | JumonjiC domain-containing histone demethylase |
| JMJC | Jumonji C |
| LSD1 | Lysine specific demethylase 1 |
| MEF | Mouse embryonic fibroblast |
| PcG | Polycomb group |
| polII | RNA polymerase II |
| PRC1 | Polycomb repressive complex 1 |
| PRE | Polycomb repressive element |

| | |
|---|---|
| qPCR | Quantitative polymerase chain reaction |
| RBAP48 | Retinoblastoma associated protein 48 |
| RNAP | RNA polymerase |
| RT | Reverse transcription |
| SAM | s-adenosyl methioine |
| SET | Su(var)-E(z)-Trithorax |
| shRNA | Short-hairpin RNA |
| SUZ12 | Suppressor of zeste 12 |
| trxG | Trithorax group |
| uH2A | Histone H2A ubiquitylated at lysine 119 |
| uH2B | Histone H2B ubiquitylated at lysine 120 |
| XCI | X chromosome inactivation |

# Chapter 1

# Introduction

Epigenetics is defined as the stable and heritable control of gene expression without regards to primary DNA sequence. This control is brought about by several broad classes of events which can alter chromatin structure on a local or global scale. These alterations in chromatin structure include the remodeling of local chromatin through ATP-dependent nucleosome repositioning, the incorporation of histone variants into the nucleosome core particle, and the covalent modification of DNA or histones. It is becoming increasingly clear that the combination of these epigenetic events at genomic loci have a profound effect on gene transcription and overall cellular physiology (summarized in (Berger, Kouzarides et al. 2009).

**THE BEGINNINGS OF A FIELD**

Well over one hundred years ago, Frederick Miescher was the first to describe the DNA/protein complexes which were housed within the nucleus of salmon sperm. Decades before the role of chromosomes in propagating genetic information was recognized, biochemists began purifying and characterizing the small basic protein component (protamine) of what Miescher named nuclein. A German biochemist, Albrecht Kossel, expanded upon Miescher's preliminary discovery and isolated a similar material from chicken blood. He noticed that the basic proteins associated with the biological acids were distinct from Miescher's protamine and named the protein histone. Over the course of the next fifty years, extensive characterization of these DNA associated proteins was undertaken with a primary focus on purification procedures and comparison between organisms and cell types. The major reason for these early studies was quite simple. Namely, the proteins were so abundant in available cellular preparations that enough starting material could be obtained for the biochemical methods available at the time. Given the ease of sample preparation,

researchers began using the DNA associated proteins to hone techniques utilized to understand the fundamentals of protein chemistry and biology en masse (Various 1971).

Throughout these early studies, biochemists believed that the protein found tightly bound to DNA was homogenous. It was not until 1950 that it was shown that this was in fact untrue and the widely studied protein preparations were heterogeneous mixtures (Stedman 1950). This original work separated histones into two groups which consisted of the lysine rich and arginine rich histones and today we know these pools as histone H1 and histone H3/H4, respectively (Kornberg and Lorch 1999). This finding ignited activity in the field to focus on potential variation between histone pools among diverse organisms and cell types, for scientists believed that histone variation throughout cellular lineages could explain the drastic differences in cellular physiology present in multicellular organisms. However, this explosion of research resulted in conflicting findings that have since been attributed to protein degradation, fraction heterogeneity, and protein aggregate formation (Various 1971). We now understand that the five main classes of histone proteins, H1, H2A, H2B, H3, and H4, demonstrate striking conservation throughout evolution. However, Stedman and Stedman's prediction of the importance of histone proteins on cellular identity and multicellular organism development lies at the very heart of the field of epigenetics to this day.

**CHROMATIN AND THE CENTRAL DOGMA**

Given a number of reports that provided circumstantial evidence regarding nuclear localization of the genetic factors, researchers focused attention on this sub-cellular compartment. This burst of investigation led to the accumulation of evidence which prompted Francis Crick to sketch out his Central Dogma in 1958. In his work, he clearly

stated that the decoding of genetic information was directional and processive with the majority of information flowing from DNA, through an RNA intermediate, into protein (Crick 1970). The remainder of this section will focus on early work in histone biology and biochemistry and how this fit into the Central Dogma.

Crick's seminal work in molecular biology paid tribute to almost forty years of scientific discovery which spanned from the work of Morgan and colleagues with *Drosophila* genetics to the derivation of DNA macromolecular structure and the purification of the enzymes capable of DNA replication and RNA polymerization. As all of these studies began falling into place, the histone biochemists remained hard at work characterizing their abundant protein which was intimately linked with the genetic information. Among the groups characterizing the properties of histone was that of Stedman and Stedman. Their work was a direct result of advances in biochemical techniques which allowed for a higher resolution separation of extracted histones coupled to powerful assays that allowed for the quantification of the relative abundance of key amino acids within these fractions. Using this technique they separated histone preparations into two pools which they named the arginine rich and lysine rich fractions, a characterization which stimulated many pioneers in the field of histone biochemistry to focus on histone fraction heterogeneity throughout diverse organisms and cell types. Interestingly, it is now clear that histones are remarkably conserved throughout eukaryotic life and while these studies may be viewed as distracting, the derivation of separation techniques and reconstitution experiments represent fundamental advances that still firmly reside in the experimental arsenal of biochemists to this day.

Between the years of 1960 and 1961 two additional pieces of the Central Dogma puzzle were put into place. First, the purification of DNA dependent RNA polymerase was

described (reviewed in (Hurwitz 2005)). Second, the biological intermediate molecule which connected DNA to protein synthesis was proven to be mRNA (Brenner, Jacob et al. 1961) retiring the one gene one ribosome model of molecular biology for good. Histones had long been known to inhibit a wide range of enzymatic activities *in vitro*. However, the purification of RNAP along with a sensitive method to detect polymerase activity provided the histone biochemist with the tools necessary to prove the long standing hypothesis that histones acted as gene repressors. Seminal work from Huang and Bonner described a method whereby chromatin could be reconstituted *in vitro* from its component histones and DNA. The authors discussed the use of DNA melting curve analysis in monitoring the reconstitution experiment and this method which employs step-wise salt dilution is still in use to this day (Huang and Bonner 1962). When chromatin which had been reconstituted with purified pea embryo histones was used in the *in vitro* RNAP assay, researchers found that it significantly repressed RNA synthesis (Huang and Bonner 1962). Trypsin is very efficient at cleaving arginine and lysine peptide bonds, two amino acids that are prevalent in histones. This fact armed researchers with the tools necessary to confirm *in vitro* results within a cellular system. Specifically, incubation of cells with trypsin resulted in the degradation of histones *in vivo* and led to a significant increase in RNA synthesis (Allfrey, Littau et al. 1963).

Advances in biochemical techniques which allowed for the automated identification of protein amino acid composition came about in 1958 (Moore, Spackman et al. 1958) and were quickly put to work on histone extracts from a diverse range of cell types. In addition to confirming the heterogeneity of original histone preparations, they also revealed absorbance peaks which could not be attributed to the standard amino acids (Rasmussen,

Murray et al. 1962). Through comparison of histone fractions with proteins containing post-translationally modified lysines, a small fraction of both methyl-lysine and acetyl-lysine residues were found within the histone polypeptides (Phillips 1963; Murray 1964). Importantly, chromatin reconstituted with acetyl-lysine histones could not inhibit RNA synthesis *in vitro* (Allfrey, Faulkner et al. 1964). The authors comment in this work,

> The findings introduce the possibility that histone effects on nuclear RNA metabolism may involve more than a simple inhibition of RNA synthesis, and that more subtle mechanisms may exist which permit both inhibition and reactivation of RNA production at different loci along the chromosome (Allfrey, Faulkner et al. 1964).

Even though the presence of modified lysine residues had been detected, it took an additional five years for experiments to show that the methylation of histone H4 was site specific *in vivo*. Careful fractionation and analysis of tryptic peptides allowed DeLange and colleagues to deduce the complete amino acid sequence of H4 including methyl-lysine at position 20 (H4K20) (DeLange, Fambrough et al. 1969). This work was followed up by the complete sequence of histone H3 which included the specific methylated lysines at position 9 and 27 (H3K9 and H3K27, respectively) (DeLange, Hooper et al. 1973) and the identification of H3 methylated at lysine 4 (H3K4) (Honda, Dixon et al. 1975).

Further studies focused on the post-translational modification of the histone proteins have revealed amino acid phosphorylation (Ord and Stocken 1966), sumoylation (Paik and Kim 1970; Shiio and Eisenman 2003), glycosylation and ADP-ribosylation (Levy-Wilson 1983), as well as arginine methylation (Paik and Kim 1970). Perhaps the most striking histone modification based on size and projected effect on overall histone topology is the addition of a single molecule of ubiquitin to histones H2A and H2B (Various 1988). Ubiquitin is a 76 amino acid protein present in all eukaryotic cells which was originally

identified based on its ability to stimulate lymphocyte differentiation *in vitro* (Goldstein, Scheid et al. 1975). Since this initial characterization, the protein has been found covalently linked to numerous cellular proteins in both monomeric and oligomeric forms. While oligomeric addition of ubiquitin to polypeptides is involved in targeting the protein for degradation (Ciechanover 1994), the monomeric addition has been implicated in numerous functional outcomes (Hicke 2001). The first identified instance of ubiquitin conjugation was the isopeptide linkage of a single ubiquitin molecule to histone H2A at lysine 119 (Goldknopf and Busch 1977). Three years later mono-ubiquitynation of histone H2B was described (West and Bonner 1980) and its positional addition to lysine 120 was identified several years later (Thorne, Sautiere et al. 1987).

In addition to the identification of modified lysines, the amino acid sequence studies described above established secondary structure predictions of the individual histones based on similarities to known globular proteins. Researchers noticed that only the carboxy-terminal portion of histones H3 and H4 had properties reminiscent of globular proteins and while the central portion of histones H2A and H2B had globular characteristics, their termini did not (Various 1971). This hypothesis has since been validated using tryptic digest (Weintraub and Van Lente 1974), and more convincingly, crystallographic techniques (Luger, Mader et al. 1997). Interestingly, all of the modified amino acids discussed to this point are located within the unstructured regions of the histones. However, the biological function of histone methylation and the effect the site-specific mark had over DNA dependent processes was not fully realized for many years to come (see below).

**CHROMATIN STRUCTURE**

If the 1960s fueled an explosion of research into the heterogeneity of histone fractions from the decades prior, the most remarkable advances in the fledgling field of epigenetics and chromatin biology of the 1970s were focused on structure-function relationships between chromatin and cellular enzymology. In 1974, the first electron micrographs of intact chromatin revealed a 10 nm fiber that was termed "particles on a string" by the authors (Olins and Olins 1974). This was promptly recoined as "beads on a string" later this same year when the elegant biochemical analyses of Roger Kornberg and colleagues were published. This work described the oligomeric nature of histone particles as containing two copies each of the four core histones (Kornberg and Thomas 1974) and the repetitive nature of histone oligomer association with DNA (Kornberg 1974). The effect of histone H1 on this structure was described a few years later (Thoma and Koller 1977) when researchers visualized chromatin in the presence and absence of the linker histone and found that H1 incorporation resulted in a 30 nm fiber. Single crystal x-ray diffraction of the histone core particle at 3.3 Å confirmed its octameric status (Burlingame, Love et al. 1985) and higher resolution structural analysis of the nucleosome, defined as the histone core particle associated with DNA, revealed the superhelical association of 146 base pairs wrapped around the octamer core (Luger, Mader et al. 1997).

Thus far, this introduction has focused on the foundation of the field of chromatin modification and epigenetic regulation of gene transcription. By the mid-1990's, scientists had amassed enough circumstantial evidence to accurately predict that alterations of local chromatin structure could have profound effects on gene transcription. Furthermore, the implication of this control on cellular differentiation and cancer were well recognized. It was clear that the nucleosome represented the repeating unit of chromatin, and specific amino

acids located along the unstructured region within histone polypeptides were subject to post-translational modification. Finally, a clear link had been described between histone tail acetylation and active gene transcription. However, the enzymes responsible for the *in vivo* modification of histone tails, as well as the role of other histone modifications in transcriptional control remained largely undetermined.

## MODERN CHROMATIN BIOLOGY

The discoveries of the first histone acetyltransferase (HAT) (Brownell and Allis 1995) and histone deacetylase (HDAC) (Taunton, Hassig et al. 1996) capable of directly acting on histone tails welcomed a new age in the functional characterization of chromatin and the modifications it harbors. Since these initial discoveries, the enzymes responsible for catalyzing both the addition and removal of identified histone modifications have been purified. In many cases, genetic disruption of these enzymes has revealed the functional implication of a given histone tail modification. Given that the scope of this work is focused primarily on histone lysine methylation and ubiquitylation, only these modifications will be discussed in detail in the following section. For more information regarding histone acetylation, arginine methylation, phosphorylation, sumoylation and glycosylation the author would like to direct the reader to several in depth reviews and the references contained therein (Wysocka, Allis et al. 2006; Oki, Aihara et al. 2007).

### *Histone lysine methylation*

As discussed in previous sections, several histone lysine residues can be methylated including K4, K9, K27, K36, and K79 of H3 and K20 of H4 (Martin and Zhang 2005). Each of these sites is located within the unstructured amino terminal tail of the histones with the exception of H3K79 which is located within the globular domain (Feng, Wang et al. 2002).

To complicate the issue further, a particular lysine residue can accept one, two, and three individual methyl groups. A variety of genetic and biochemical techniques have been used to correlate methylated lysine residues to transcriptional competence. The first example of this was the demonstration that H3K4me in *Tetrahymena* and yeast was associated with actively transcribed macronuclei and acetylated histones, respectively (Strahl, Ohba et al. 1999). Additional methyl marks associated with gene activation include H3K79me and H3K36me. H3K9me was demonstrated to be essential for eukaryotic heterochromatin formation (Bannister, Zegerman et al. 2001) and thus correlated to gene repression. In addition, H3K27me and H4K20me have also been implicated in gene silencing (Martin and Zhang 2005).

The first enzyme capable of histone methyltransferase activity to be purified was the H3K9 specific HMT, SUV39H1, which was found to be the mammalian homolog of fission yeast Clr4 and *Drosophila* suppressor of position effect variegation Su(var)3-9 (Rea, Eisenhaber et al. 2000). Importantly, the active site of this enzyme was mapped to the Su(var)-E(Z)-Trithorax (SET) domain which is evolutionarily conserved throughout all eukaryotes and is present in multiple proteins in individual species (Dillon, Zhang et al. 2005). In addition to SUV39H1, at least 15 additional mammalian HMTs have been identified (Martin and Zhang 2005) with activity directed against various lysine residues and the variable capacity to methylate to all three states. These additional enzymes include the mammalian homolog of yeast Set2 (Strahl, Grant et al. 2002), Setd2 (Sun, Wei et al. 2005), a H3K36 specific HMT and Ezh2, a H3K27 specific HMT (Cao, Wang et al. 2002). Of all the lysine HMTs identified, all possess the SET domain with the exception of Dot1l, an H3K79 specific HMT conserved from yeast to human (Feng, Wang et al. 2002).

In addition to the general correlations presented above, HMT identification has led to further insight into the functional role that histone modifications play throughout the life-cycle of a cell. For example, fission yeast protein Set2 has been found to associate with the phosphorylated form of elongating RNA Polymerase II (polII) (Morris, Shibata et al. 2005). This association leads to the enrichment of the H3K36me mark within the coding regions of expressed genes (Rao, Shibata et al. 2005) and plays an important role in suppressing polII transcription from cryptic start sites (Li, Gogol et al. 2007; Li, Jackson et al. 2009). While these findings in yeast have yet to be extrapolated to mammalian systems, localization of the H3K36me mark in higher eukaryotes has revealed a similar enrichment of the modification within the transcribed region of expressed genes (Bell, Wirbelauer et al. 2007; Mikkelsen, Ku et al. 2007).

The first example of active removal of methyl groups from histone tail lysine residues came from the purification and characterization of the Lysine Specific Demethylase 1 (LSD1) (Shi, Lan et al. 2004). This enzyme has been found to remove methyl marks from H3K4 and H3K9 through an oxidative mechanism which utilizes a flavin cofactor (Forneris, Binda et al. 2005). Characterization of LSD1 has led to the discovery of the role H3K9me plays in inducible gene silencing. Specifically, the enzyme is recruited to Androgen Receptor (AR) target genes where the removal of repressive H3K9me is necessary for the efficient induction of gene expression (Metzger, Wissmann et al. 2005).

As discussed above, the diversity and scope of histone methylation is vast and the discovery of LSD1 raised two additional questions regarding the turn-over of histone methylation. First, the chemistry fueling the oxidative reaction could only be applied to mono- and di-methylated lysine residues. Second, phylogenetic analysis could only identify

one additional paralogue in mammalian cells. These two enzymes could not account for active removal of methyl groups from all sites of histone modification (Culhane and Cole 2007). These two facts prompted researchers to continue biochemical screens for enzymes capable of active histone demethylation and a H3K36 specific demethylase, Jumonji containing Histone Demethylase 1 (Jhdm1), was identified (Tsukada, Fang et al. 2006). Enzymatic activity was mapped to the Jumonji C (JmjC) domain and was shown to adhere to a two-step hydroxylation reaction which utilized Fe(II) and α-ketoglutarate as cofactors (Klose, Kallin et al. 2006). Similar to the SET domain family of HMTs, numerous mammalian proteins contained a JmjC domain and it was recognized that this protein super-family could contain demethylases specific for additional histone residues. To date, JmjC domain-containing demethylases have been identified that are capable of removing H3K4me1/2/3, H3K9me2/3, H3K27me2/3, and H3K36me1/2/3 ((Klose, Yamane et al. 2006; Agger, Cloos et al. 2007; Iwase, Lan et al. 2007; Klose, Yan et al. 2007; Yamane, Tateishi et al. 2007)). Characterization of these enzymes has revealed their role, as well as the role of the methylations they erase, in cellular processes which include embryonic stem cell pluripotency, X chromosome inactivation (XCI), and hormone receptor signaling pathways (reviewed in (Takeuchi, Watanabe et al. 2006; Benevolenskaya 2007; Swigut and Wysocka 2007)).

*Histone Ubiquitylation*

Experimental evidence suggests that the addition of ubiquitin to lysine 119 of H2A and lysine 120 of H2B result in very different functional outcomes. The majority of these findings rely on the isolation and characterization of the enzymes required for the placement and removal of the ubiquitin mark. The first enzyme identified as responsible for histone

ubiquitylation was budding yeast protein Rad6, a uH2B specific ligase that had previously been linked to DNA repair (Robzyk, Recht et al. 2000). Genetic manipulation of Rad6 activity confirmed a general association of the uH2B mark with active chromatin (Davie and Murphy 1990) and revealed that Rad6-dependent ubiquitylation of H2B was necessary for the enrichment of H3K79me and H3K4me, two additional histone marks intimately linked to active gene expression (Lee, Shukla et al. 2007). Since this initial discovery, an elegant biochemical approach has revealed that uH2B allosterically stimulates Dot1 activity to methylate H3K79 (McGinty, Kim et al. 2008).

uH2A has been historically linked to gene expression based on classic two-dimensional separation studies which demonstrated that it co-migrated with several actively transcribed loci on native gels (Barsoum and Varshavsky 1985). In light of these early studies it is surprising that the first identified enzyme capable of placing this mark, Ring1b, was a component of the Polycomb repressive Complex 1 (PRC1), which has been implicated in the establishment and maintenance of gene repression (Wang, Wang et al. 2004) (see below). Additional evidence for a role of uH2A in gene silencing has stemmed from the identification of 2A-HUB, another H2A targeted ubiquitin ligase, and the characterization of four separate H2A specific deubiquitylases (reviewed in (Zhou, Wang et al. 2009)).

*Polycomb-mediated gene repression*

An emerging paradigm in epigenetic regulation of cellular physiology is represented by the opposing effects of Polycomb Group (PcG) and Trithorax Group (trxG) proteins in the maintenance of homeotic transcription factor (Hox) gene transcription during *Drosophila* development (Ringrose and Paro 2004). The spatial and temporal expression pattern of Hox genes is controlled through silencing by PcG and activation by trxG. Originally identified

through classic fly genetics, molecular biology and modern chromatin enzymology have linked these genes to factors involved in directly altering local chromatin structure and have proven them conserved in higher eukaryotes (Cao and Zhang 2004). These studies have associated many of the trxG group genes to transcription permissible chromatin modifications such as histone acetylation and H3K4me. Alternatively, PcG genes have been implicated in catalyzing histone H3K27me and uH2A as described in detail below.

Arguably the greatest advance in the study of PcG mediated gene repression has been the biochemical purification of two protein complexes, PRC1 and PRC2, with distinct chromatin modifying capabilities. The study of these complexes, and the individual components they harbor, has led to unprecedented understanding of eukaryotic gene repression and facultative heterochromatin formation. The catalytic core of the human PRC1 complex is composed of RING1A, RING1B, BMI1, and HPH2 (Wang, Wang et al. 2004) and has been extensively studied. *In vitro* reconstitution experiments and *in vivo* loss of function studies have demonstrated that Ring1b is responsible for the vast majority of uH2A in cells. Furthermore, both Ring1a and Bmi1can stimulate the ligase activity of Ring1b *in vitro* and *in vivo.* Disruption of any of these proteins *in vivo* leads to the targeted loss of uH2A and aberrant HOX gene expression (Cao, Tsukada et al. 2005; Buchwald, van der Stoop et al. 2006). Interestingly, mammalian cells contain at least three Bmi1 paralogues which have been shown to substitute for the protein in PRC1 under various cellular contexts (Elderkin, Maertens et al. 2007; Wu, Gong et al. 2008). The mammalian PRC2 complex, whose catalytic core is comprised of Ezh2, Suz12, Eed, Aebp2, and RbAp48, is responsible for the majority of H3K27me in eukaryotic cells (Cao, Wang et al. 2002) and Ezh2, Suz12, and Eed are absolutely essential for H3K27me *in vitro* (Cao and Zhang 2004). Furthermore,

aberrant HOX gene expression upon loss of Ezh2 function, as well as its fly counterpart E(z), is accompanied by loss of H3K27me at these promoters (Cao, Wang et al. 2002).

The mechanism driving chromatin modifying enzyme targeting represents a major question in modern transcription research. While several of these enzymes harbor domains with demonstrated DNA binding potential (Martin and Zhang 2005), primary DNA sequence can not be the only deciding factor in their recruitment. First, the average DNA binding domain can recognize motifs which exist tens of thousands of times in higher eukaryotic genomes (Hannenhalli 2008). Furthermore, chromatin modifiers are differentially targeted in any number of cellular lineages which contain identical DNA sequence (Ku, Koche et al. 2008). The study of Polycomb mediated gene silencing has provided valuable insight into the mechanism of complex targeting in *Drosophila* as directed by defined Polycomb Responsive Elements (PREs) (reviewed in (Schwartz and Pirrotta 2008)). These genomic elements can coordinate the binding of the PRC2 through sequence specific interactions with DNA binding proteins such as PHO/YY1 (Mohd-Sarip, van der Knaap et al. 2006). However, a mechanism for selective PRC2 binding throughout cellular lineages, as well as homologous elements in mammalian cells, has yet to be described.

In addition to extending our knowledge regarding chromatin modification and transcriptional competence, studies of PcG mediated silencing have also highlighted the role of specific histone modifications in the recruitment of additional factors which contain distinct protein domains. These domains are highly conserved throughout evolution and are present in a variety of transcriptional coactivators, corepressors, and chromatin structural components (reviewed in (Daniel, Pray-Grant et al. 2005; Zhang 2006)). Specific binding of acetylated lysine residues has been attributed to bromodomains, while lysine binding at

various methylation states and residues has been linked to chromo, PHD, WD-40 and Tudor domains (Zhang 2006). In addition to the catalytic core described above, the PRC1 complex also contains component Pc which possesses a chromodomain with H3K27me3 binding potential. This association has established a molecular mechanism whereby the direct result of PRC2 HMT activity can recruit PRC1dependent ubiquitin ligase activity. Interestingly, this mechanistic insight can explain the genetic interaction of these two distinct protein complexes (Cao, Wang et al. 2002; Fischle, Wang et al. 2003).

While it is impossible to understate the importance of PcG proteins in multicellular organism development, both PRC1 and PRC2 have been linked to a variety of biological functions that are not limited to HOX gene regulation. For example, PRC2 mediated gene repression is crucial for the establishment and maintenance of XCI (Plath, Fang et al. 2003), the maintenance of embryonic stem cell pluripotency (Pasini, Bracken et al. 2007), and the silencing of key tumor suppressor genes in mammalian cells (Bracken, Kleine-Kohlbrecher et al. 2007; Kotake, Cao et al. 2007). Additionally, PRC1 has been implicated in XCI (Schoeftner, Sengupta et al. 2006) as well as tumor suppressor silencing (Kotake, Cao et al. 2007). Knock-out of PRC1 complex components results in the upregulation of $p15^{Ink4a}$ and the activation of a pRB mediated cellular senescence program (Kotake, Cao et al. 2007).

*The genomics age*

Recent technological advances have opened the door to genome wide localization studies of chromatin remodeling factors and the covalent modifications they enact (Schones and Zhao 2008). Chromatin immunoprecipitation coupled to genomic microarrays (ChIP-chip) or high throughput sequencing (ChIP-seq) have led to genome wide identification of factor binding sites and histone modifications in numerous cell types that span organism

phylogeny. These powerful techniques have enabled the extrapolation of what may be individual target gene regulation events to the definition of general trends in regulatory mechanisms and are proving to be instrumental in the mechanistic dissection of complex factor targeting in higher eukaryotes (Mendenhall and Bernstein 2008).

Over one hundred years of scientific experimentation has fueled the evolution of our understanding regarding the importance of chromatin in gene transcription and the regulation of virtually all fundamental biological processes. The identification of modified histones, the elucidation of the enzymes which catalyze covalent linkage, and the identification of distinct protein domains which possess modification state binding properties has led to formation of the histone code hypothesis (Strahl and Allis 2000). This posits that the combination of covalent modifications within a given gene regulatory region can dictate its transcriptional competence. Furthermore, regulation is the result of both intrinsic nucleosome stability, as well as additional effector proteins which can recognize and bind to a given cohort of marks. Most importantly, the hypothesis suggests that a clear understanding of the interplay between all of these factors could allow for transcriptional manipulation of biological systems. This manipulation would have a dramatic impact on our ability to understand and treat a multitude of human diseases.

**RESEARCH SUMMARY**

My dissertation highlights the use of four separate analytical methods that are commonly employed in chromatin research: phylogenetic analysis, biochemistry, molecular genetics and genome-wide localization. These technical categories are used to examine three chromatin related questions. First, can the conservation of JmjC domains reveal distinct classes of histone demethylases as defined by evolutionary conservation and maintenance of

amino acid residues critical for activity? Second, what is the biochemical activity and molecular function of a specific histone demethylase, Jhdm1b, in primary cells? Third, what is the contribution of PRC1 component Bmi1 to uH2A deposition genome-wide?

At the commencement of my first project, two pieces of critical information were known. First, the biochemical purification of two histone demethylases linked enzymatic activity to the JmjC domain (Tsukada, Fang et al. 2006; Yamane, Toumazou et al. 2006). Second, the crystal structure of FIH, a JmjC domain-containing protein implicated in the hydroxylation of asparagine residues, had been solved in the presence of important co-factors which stimulate activity (Lee, Kim et al. 2003). This information, coupled with extensive database curation, was used to reveal the evolutionary conservation of the JmjC domain-containing proteins and to predict their ability to catalyze the direct removal of methyl groups from histone protein substrates.

At the time of my second project, our lab had characterized the H3K36me substrate specificity of Jhdm1a *in vitro* and *in vivo*. In addition, the phylogenetic studies described above had demonstrated extensive JmjC domain homology with Jhdm1b, its mammalian paralogue. However, another report had implicated Jhmd1b in directing methyl group removal from H3K4 *in vivo* (Frescas, Guardavaccaro et al. 2007). In addition, conflicting screens had demonstrated both oncogenic and tumor suppressive potential for Jhdm1b (Suzuki, Minehata et al. 2006; Pfau, Tzatsos et al. 2008). We set out to settle both of these apparent discrepancies using a combination of biochemistry and molecular biology in primary cells.

At the time my third project began, the power of ChIP-seq technology had been described in a series of studies defining the genome-wide chromatin modification patterns in

cells. However, this technique had never been applied to uH2A. In addition, the functional outcome of Bmi1 dependent uH2A deposition at several target genes had implicated the mark in transcriptional repression. I applied genome-wide studies to define the portion of the genome associated with Bmi1-dependent uH2A and to investigate whether the results of individual gene studies could be extrapolated genome-wide.

## REFERENCES


Agger, K., P. A. Cloos, et al. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**(7163): 731-4.

Allfrey, V. G., R. Faulkner, et al. (1964). Acetylation And Methylation Of Histones And Their Possible Role In The Regulation Of Rna Synthesis. *Proc Natl Acad Sci U S A* **51**: 786-94.

Allfrey, V. G., V. C. Littau, et al. (1963). On the role of of histones in regulation ribonucleic acid synthesis in the cell nucleus. *Proc Natl Acad Sci U S A* **49**: 414-21.

Bannister, A. J., P. Zegerman, et al. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**(6824): 120-4.

Barsoum, J. and A. Varshavsky (1985). Preferential localization of variant nucleosomes near the 5'-end of the mouse dihydrofolate reductase gene. *J Biol Chem* **260**(12): 7688-97.

Bell, O., C. Wirbelauer, et al. (2007). Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila. *Embo J* **26**(24): 4974-84.

Benevolenskaya, E. V. (2007). Histone H3K4 demethylases are essential in development and differentiation. *Biochem Cell Biol* **85**(4): 435-43.

Berger, S. L., T. Kouzarides, et al. (2009). An operational definition of epigenetics. *Genes Dev* **23**(7): 781-3.

Bracken, A. P., D. Kleine-Kohlbrecher, et al. (2007). The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes Dev* **21**(5): 525-30.

Brenner, S., F. Jacob, et al. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190**(4776): 576-581.

Brownell, J. E. and C. D. Allis (1995). An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. *Proc Natl Acad Sci U S A* **92**(14): 6364-8.

Buchwald, G., P. van der Stoop, et al. (2006). Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. *Embo J* **25**(11): 2465-74.

Burlingame, R. W., W. E. Love, et al. (1985). Crystallographic structure of the octameric histone core of the nucleosome at a resolution of 3.3 A. *Science* **228**(4699): 546-53.

Cao, R., Y. Tsukada, et al. (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol Cell* **20**(6): 845-54.

Cao, R., L. Wang, et al. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**(5595): 1039-43.

Cao, R. and Y. Zhang (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr Opin Genet Dev* **14**(2): 155-64.

Cao, R. and Y. Zhang (2004). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell* **15**(1): 57-67.

Ciechanover, A. (1994). The ubiquitin-proteasome proteolytic pathway. *Cell* **79**(1): 13-21.

Crick, F. (1970). Central dogma of molecular biology. *Nature* **227**(5258): 561-3.

Culhane, J. C. and P. A. Cole (2007). LSD1 and the chemistry of histone demethylation. *Curr Opin Chem Biol* **11**(5): 561-8.

Daniel, J. A., M. G. Pray-Grant, et al. (2005). Effector proteins for methylated histones: an expanding family. *Cell Cycle* **4**(7): 919-26.

Davie, J. R. and L. C. Murphy (1990). Level of ubiquitinated histone H2B in chromatin is coupled to ongoing transcription. *Biochemistry* **29**(20): 4752-7.

DeLange, R. J., D. M. Fambrough, et al. (1969). Calf and pea histone IV. II. The complete amino acid sequence of calf thymus histone IV; presence of epsilon-N-acetyllysine. *J Biol Chem* **244**(2): 319-34.

DeLange, R. J., J. A. Hooper, et al. (1973). Histone 3. 3. Sequence studies on the cyanogen bromide peptides; complete amino acid sequence of calf thymus histone 3. *J Biol Chem* **248**(9): 3261-74.

Dillon, S. C., X. Zhang, et al. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol* **6**(8): 227.

Elderkin, S., G. N. Maertens, et al. (2007). A phosphorylated form of Mel-18 targets the Ring1B histone H2A ubiquitin ligase to chromatin. *Mol Cell* **28**(1): 107-20.

Feng, Q., H. Wang, et al. (2002). Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr Biol* **12**(12): 1052-8.

Fischle, W., Y. Wang, et al. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* **17**(15): 1870-81.

Forneris, F., C. Binda, et al. (2005). Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. *FEBS Lett* **579**(10): 2203-7.

Frescas, D., D. Guardavaccaro, et al. (2007). JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes. *Nature* **450**(7167): 309-13.

Goldknopf, I. L. and H. Busch (1977). Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24. *Proc Natl Acad Sci U S A* **74**(3): 864-8.

Goldstein, G., M. Scheid, et al. (1975). Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *Proc Natl Acad Sci U S A* **72**(1): 11-5.

Hannenhalli, S. (2008). Eukaryotic transcription factor binding sites--modeling and integrative search methods. *Bioinformatics* **24**(11): 1325-31.

Hicke, L. (2001). Protein regulation by monoubiquitin. *Nat Rev Mol Cell Biol* **2**(3): 195-201.

Honda, B. M., G. H. Dixon, et al. (1975). Sites of in vivo histone methylation in developing trout testis. *J Biol Chem* **250**(22): 8681-5.

Huang, R. C. and J. Bonner (1962). Histone, a suppressor of chromosomal RNA synthesis. *Proc Natl Acad Sci U S A* **48**: 1216-22.

Hurwitz, J. (2005). The discovery of RNA polymerase. *J Biol Chem* **280**(52): 42477-85.

Iwase, S., F. Lan, et al. (2007). The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell* **128**(6): 1077-88.

Klose, R. J., E. M. Kallin, et al. (2006). JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet* **7**(9): 715-27.

Klose, R. J., K. Yamane, et al. (2006). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* **442**(7100): 312-6.

Klose, R. J., Q. Yan, et al. (2007). The retinoblastoma binding protein RBP2 is an H3K4 demethylase. *Cell* **128**(5): 889-900.

Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science* **184**(139): 868-71.

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

Kornberg, R. D. and J. O. Thomas (1974). Chromatin structure; oligomers of the histones. *Science* **184**(139): 865-8.

Kotake, Y., R. Cao, et al. (2007). pRB family proteins are required for H3K27 trimethylation and Polycomb repression complexes binding to and silencing p16INK4alpha tumor suppressor gene. *Genes Dev* **21**(1): 49-54.

Ku, M., R. P. Koche, et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**(10): e1000242.

Lee, C., S. J. Kim, et al. (2003). Structure of human FIH-1 reveals a unique active site pocket and interaction sites for HIF-1 and von Hippel-Lindau. *J Biol Chem* **278**(9): 7558-63.

Lee, J. S., A. Shukla, et al. (2007). Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell* **131**(6): 1084-96.

Levy-Wilson, B. (1983). Glycosylation, ADP-ribosylation, and methylation of Tetrahymena histones. *Biochemistry* **22**(2): 484-9.

Li, B., M. Gogol, et al. (2007). Infrequently transcribed long genes depend on the Set2/Rpd3S pathway for accurate transcription. *Genes Dev* **21**(11): 1422-30.

Li, B., J. Jackson, et al. (2009). Histone H3 Lysine 36 Dimethylation (H3K36me2) Is Sufficient to Recruit the Rpd3s Histone Deacetylase Complex and to Repress Spurious Transcription. *J Biol Chem* **284**(12): 7970-6.

Luger, K., A. W. Mader, et al. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**(6648): 251-60.

Martin, C. and Y. Zhang (2005). The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* **6**(11): 838-49.

McGinty, R. K., J. Kim, et al. (2008). Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **453**(7196): 812-6.

Mendenhall, E. M. and B. E. Bernstein (2008). Chromatin state maps: new technologies, new insights. *Curr Opin Genet Dev* **18**(2): 109-15.

Metzger, E., M. Wissmann, et al. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**(7057): 436-9.

Mikkelsen, T. S., M. Ku, et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**(7153): 553-60.

Mohd-Sarip, A., J. A. van der Knaap, et al. (2006). Architecture of a polycomb nucleoprotein complex. *Mol Cell* **24**(1): 91-100.

Moore, S., D. H. Spackman, et al. (1958). Automatic recording apparatus for use in the chromatography of amino acids. *Fed Proc* **17**(4): 1107-15.

Morris, S. A., Y. Shibata, et al. (2005). Histone H3 K36 methylation is associated with transcription elongation in Schizosaccharomyces pombe. *Eukaryot Cell* **4**(8): 1446-54.

Murray, K. (1964). The Occurrence Of Epsilon-N-Methyl Lysine In Histones. *Biochemistry* **3**: 10-5.

Oki, M., H. Aihara, et al. (2007). Role of histone phosphorylation in chromatin dynamics and its implications in diseases. *Subcell Biochem* **41**: 319-36.

Olins, A. L. and D. E. Olins (1974). Spheroid chromatin units (v bodies). *Science* **183**(4122): 330-2.

Ord, M. G. and L. A. Stocken (1966). Metabolic properties of histones from rat liver and thymus gland. *Biochem J* **98**(3): 888-97.

Paik, W. K. and S. Kim (1970). Omega-N-methylarginine in histones. *Biochem Biophys Res Commun* **40**(1): 224-9.

Pasini, D., A. P. Bracken, et al. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* **27**(10): 3769-79.

Pfau, R., A. Tzatsos, et al. (2008). Members of a family of JmjC domain-containing oncoproteins immortalize embryonic fibroblasts via a JmjC domain-dependent process. *Proc Natl Acad Sci U S A* **105**(6): 1907-12.

Phillips, D. M. (1963). The presence of acetyl groups of histones. *Biochem J* **87**: 258-63.

Plath, K., J. Fang, et al. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**(5616): 131-5.

Rao, B., Y. Shibata, et al. (2005). Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Mol Cell Biol* **25**(21): 9447-59.

Rasmussen, P. S., K. Murray, et al. (1962). On the complexity of calf thymus histone. *Biochemistry* **1**: 79-89.

Rea, S., F. Eisenhaber, et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**(6796): 593-9.

Ringrose, L. and R. Paro (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet* **38**: 413-43.

Robzyk, K., J. Recht, et al. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. *Science* **287**(5452): 501-4.

Schoeftner, S., A. K. Sengupta, et al. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *Embo J* **25**(13): 3110-22.

Schones, D. E. and K. Zhao (2008). Genome-wide approaches to studying chromatin modifications. *Nat Rev Genet* **9**(3): 179-91.

Schwartz, Y. B. and V. Pirrotta (2008). Polycomb complexes and epigenetic states. *Curr Opin Cell Biol* **20**(3): 266-73.

Shi, Y., F. Lan, et al. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**(7): 941-53.

Shiio, Y. and R. N. Eisenman (2003). Histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci U S A* **100**(23): 13225-30.

Stedman, E. (1950). Cell specificity of histones. *Nature* **166**(4227): 780-1.

Strahl, B. D. and C. D. Allis (2000). The language of covalent histone modifications. *Nature* **403**(6765): 41-5.

Strahl, B. D., P. A. Grant, et al. (2002). Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol Cell Biol* **22**(5): 1298-306.

Strahl, B. D., R. Ohba, et al. (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**(26): 14967-72.

Sun, X. J., J. Wei, et al. (2005). Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *J Biol Chem* **280**(42): 35261-71.

Suzuki, T., K. Minehata, et al. (2006). Tumor suppressor gene identification using retroviral insertional mutagenesis in Blm-deficient mice. *Embo J* **25**(14): 3422-31.

Swigut, T. and J. Wysocka (2007). H3K27 demethylases, at long last. *Cell* **131**(1): 29-32.

Takeuchi, T., Y. Watanabe, et al. (2006). Roles of jumonji and jumonji family genes in chromatin regulation and development. *Dev Dyn* **235**(9): 2449-59.

Taunton, J., C. A. Hassig, et al. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**(5260): 408-11.

Thoma, F. and T. Koller (1977). Influence of histone H1 on chromatin structure. *Cell* **12**(1): 101-7.

Thorne, A. W., P. Sautiere, et al. (1987). The structure of ubiquitinated histone H2B. *Embo J* **6**(4): 1005-10.

Tsukada, Y., J. Fang, et al. (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**(7078): 811-6.

Various (1971). *Histones and Nucleohistones*. New York, Plenum Publishing Company Ltd.

Various (1988). *Ubiquitin*. New York, Plenum Press.

Wang, H., L. Wang, et al. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**(7010): 873-8.

Weintraub, H. and F. Van Lente (1974). Dissection of chromosome structure with trypsin and nucleases. *Proc Natl Acad Sci U S A* **71**(10): 4249-53.

West, M. H. and W. M. Bonner (1980). Histone 2B can be modified by the attachment of ubiquitin. *Nucleic Acids Res* **8**(20): 4671-80.

Wu, X., Y. Gong, et al. (2008). Cooperation between EZH2, NSPc1-mediated histone H2A ubiquitination and Dnmt1 in HOX gene silencing. *Nucleic Acids Res* **36**(11): 3590-9.

Wysocka, J., C. D. Allis, et al. (2006). Histone arginine methylation and its dynamic regulation. *Front Biosci* **11**: 344-55.

Yamane, K., K. Tateishi, et al. (2007). PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Mol Cell* **25**(6): 801-12.

Yamane, K., C. Toumazou, et al. (2006). JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125**(3): 483-95.

Zhang, Y. (2006). It takes a PHD to interpret histone methylation. *Nat Struct Mol Biol* **13**(7): 572-4.

Zhou, W., X. Wang, et al. (2009). Histone H2A ubiquitination in transcriptional regulation and DNA damage repair. *Int J Biochem Cell Biol* **41**(1): 12-5.

# Chapter 2

## The evolutionarily conserved Jumonji C domain-containing proteins can be separated into seven groups based on domain architecture and phylogenetic relationship

This chapter is the result of collaboration with Dr. Robert J. Klose and portions of this text are taken from a co-authored review published on the topic (Klose, R.J., Kallin E.M., and Zhang, Y. *Nat. Rev. Genetics.* 2006). I carried out all of the described bioinformatic analysis.

**ABSTRACT**

Histone methylation plays important roles in regulating gene expression and forms part of the epigenetic memory system that regulates cell fate and identity. Recently, enzymes capable of directly removing methyl marks from histones have been identified, revealing a new level of plasticity within this epigenetic modification system. Here we analyze the evolutionary relationship between JmjC domain containing proteins with the goal of predicting additional active demethylases.

## INTRODUCTION

The basic unit of chromatin consists of 146 bp of DNA wrapped around a histone octamer which is composed of two copies of each of the four core histones H2A, H2B, H3, and H4. Covalent modification of core histones modulates genome function by contributing additional epigenetic information. One such modification is methylation, which occurs on arginine and lysine residues and is involved in regulating a wide range of chromatin based processes (reviewed in (Martin and Zhang 2005)). While arginine residues can be methylated to mono- and di- states, lysine residues can house one to three (mono, di, tri) methyl groups. The histone modification state of a given residue can have profound affects on the function of chromatin.

The identification and characterization of histone methyltransferase enzymes has provided a wealth of knowledge regarding the effects of distinct histone methylation marks on chromatin function. Until recently, it was not known whether enzymes capable of antagonizing histone methylation existed. However, during the past two years, three distinct classes of enzyme capable of antagonizing histone methylation have been characterized. PADI4 was the first of these enzymes identified and it functions as a histone deiminase that converts methyl-arginine to citrulline as opposed to directly reversing arginine methylation (Cuthbert, Daujat et al. 2004; Wang, Wysocka et al. 2004). Lysine Specific Demethylase 1 (LSD1) was the founding member of the second class of enzymes which directly reverses histone H3K4/K9 lysine modifications by an oxidative demethylation reaction requiring flavin as a co-factor (Shi, Lan et al. 2004; Metzger, Wissmann et al. 2005). The third, and largest class of demethylase enzymes, contain a JmjC domain and catalyze histone lysine demethylation via an oxidative reaction that requires iron Fe(II) and alpha-ketoglutarate

(αKG) as co-factors (Tsukada, Fang et al. 2006). Unlike LSD1, which can only remove mono and di-methyl lysine modifications, the JmjC domain containing histone demethylases are capable of removing all three histone lysine methylation states. Thus far, JmjC domain containing histone demethylases (JHDMs) have been shown to reverse H3K36 (JHMD1) (Tsukada, Fang et al. 2006), H3K9 (JHDM2A, JHDM3B/D) (Yamane, Toumazou et al. 2006) and both H3K9/36 (JHDM3A/C) methylation (Klose, Yamane et al. 2006; Whetstine, Nottke et al. 2006).

The JmjC domain was first defined based on amino acid similarities in the Jarid2 (Jumonji), Jarid1C (Smcx), and Jarid1A (RBP2) proteins (Takeuchi, Yamazaki et al. 1995; Balciunas and Ronne 2000; Clissold and Ponting 2001). Homology between the JmjC and cupin metalloenzyme domains (Dunwell and Gane 1998; Clissold and Ponting 2001) led to the identification of the JmjC domain containing factor inhibiting hypoxia (FIH) as an active protein hydroxylase (Lando, Peet et al. 2002). Given that several JmjC domain-containing proteins are known to associate with chromatin, it was proposed that some members of this protein family may carry out histone demethylation reactions using a mechanism analogous to the AlkB family of DNA demethylases (Trewick, McLaughlin et al. 2005; Tsukada, Fang et al. 2006). This prediction was confirmed through an unbiased activity-based biochemical purification which identified a JmjC domain-containing protein, JHDM1, as a H3K36-specific demethylase (Tsukada, Fang et al. 2006). Based on the crystal structure of FIH, the JmjC domain folds into eight beta sheets forming an enzymatically active pocket which coordinates Fe(II) and αKG (Dann, Bruick et al. 2002; Elkins, Hewitson et al. 2003; Lee, Kim et al. 2003). Within the JmjC domain, three amino acid residues are involved in binding to the cofactor Fe(II), and two additional residues contact with αKG. Many JmjC domain-

containing proteins have conserved residues within the predicted co-factor binding sites indicating that these proteins may be active enzymes, while other members have amino acids variations which are likely to abrogate enzymatic activity. Substrate specificity for JmjC domain containing histone demethylases appears to rely on both the JmjC domain and additional domains found within each enzyme.

Here we have analyzed JmjC domain-containing proteins from yeast, worm, fly, mouse, and human. By utilizing bioinformatic analysis we provide new insight into JmjC domain homology, protein domain architecture, and enzymatic activity. Based on our analysis, we predict that many more JmjC domain-containing proteins are likely to function as protein hydroxylases or histone demethylases.

## RESULTS

Through analysis of public protein domain databases including PFAM and SMART, we have extracted a non-redundant set of 98 JmjC domain-containing proteins from various organisms including: human (30), mouse (30), D. *melanogaster* (13), C.*elegans* (13), S.*pombe* (7), and S.*cerevisiae* (5) (Table 2-1). A combination of multiple sequence alignment (Chenna, Sugawara et al. 2003) and bayesian inference phylogeny (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) was used to visualize the evolutionary relationship between the JmjC domains of the proteins within our database (Fig. 2-1). We then divided the JmjC domain-containing proteins into seven groups by merging information regarding the domain architecture of the full length protein with JmjC domain based phylogeny (Fig. 2-1). Using this classification scheme, we have defined seven groups of evolutionarily conserved proteins, six of which have at least one additional protein domain beside the JmjC domain, and a seventh group that contains only a JmjC domain (Fig. 2-1). To further analyze the JmjC domain within individual groups, we focused on amino acid conservation within predicted Fe(II) and αKG binding sites to provide additional insight into potential enzymatic activity. To this end, all proteins that contain a JmjC domain that demonstrate conserved cofactor binding amino acids were identified (Fig. 2-1, solid triangles).

## DISCUSSION

In 2000, the SET domain was identified as the enzymatic motif shared amongst a broad class of histone methyltransferase enzymes (Rea, Eisenhaber et al. 2000). Since then there has been an exponential growth in our understanding of how histone methylation contributes to chromatin function (Martin and Zhang 2005). With the identification of enzymes that antagonize histone methylation (Cuthbert, Daujat et al. 2004; Shi, Lan et al. 2004; Wang, Wysocka et al. 2004; Tsukada, Fang et al. 2006), we appear to be on the cusp of another rapid advancement in our understanding of histone methylation.

Although it is clear that not all JmjC domain-containing proteins will be functional enzymes, our analysis suggests that many of these proteins satisfy the co-factor binding requirements and appear to be excellent candidates for histone or non-histone protein demethylases. The specific roles that individual histone or non-histone protein demethylases will play in normal cellular function remains to be revealed, but it is clear that histone methylation can be dynamically regulated in a manner analogous to acetylation or phosphorylation. The histone demethylases identified thus far have roles in counteracting histone modifications that oppose the transcription state of a given gene and in removing histone modifications during the transition from one transcriptional state to another. With the constantly expanding repertoire of site-specific histone demethylases, we look forward to determining how these unique enzymes will function in regulating transcription, epigenetic inheritance, and in controlling cell fate.

## EXPERIMENTAL PROCEDURES

*Database curation and phylogenetic tree generation:* The SMART (Schultz, Milpetz et al. 1998; Letunic, Copley et al. 2006) and the PFAM domain databases (Bateman, Coin et al. 2004) were used to obtain a list of all JmjC domain containing proteins from: human, mouse, D. *melanogaster,* C. *elegans*, fission yeast, and budding yeast. This list was then curated by extracting the chromosome location of all entries. Any JmjC domain containing proteins that were linked to the same physical chromosome location were aligned using Multalin (Corpet 1988) to eliminate any redundancy in our curated JmjC domain containing data set. Within our curated JmjC domain-containing protein database we have included: 30 human proteins, 30 mouse proteins, 13 D. *melanogastar* proteins, 13 C. *elegans* proteins, 7 fission yeast proteins and 5 budding yeast proteins. A multiple sequence alignment of the JmjC domain was performed using ClustalW (Corpet 1988; Chenna, Sugawara et al. 2003) followed by phylogenetic tree generation using a mixed amino acid model and 2.2 million generations with sampling every 1000$^{th}$ generation using Mr. Bayes (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). The resulting tree was imported into Mega 3.1 (Kumar, Tamura et al. 2004) for visualization and analysis. Based on the phylogenetic tree and the domain architecture of the full length proteins we have classified the JmjC domain containing proteins into seven groups.

**Figure 2-1.** The phylogenetic relationship between the JmjC (Jumonji C)-domain-containing proteins from model organisms. Multiple sequence alignment followed by Bayesian inference analysis was used to compute JmjC domain phylogenetic relationships in several model organisms. All posterior probabilities of clade partitions that were <100% are shown and are defined as the percentage of sampled trees used in the analysis that contained the consensus partition. Seven evolutionarily conserved groups of JmjC-domain-containing proteins were defined by combining information from the phylogenetic analysis (left) and the domain architecture of the full-length protein (right). The asterisk indicates proteins that were placed in a given group based on homology within the JmjC domain, but that lack some aspects of the domain architecture found in their related orthologues. The JmjC-domain-only group has been divided into eight subgroups based on similarity within the JmjC domain, as determined by phylogenetic analysis alone. Schizosaccharomyces pombe (sp); Caenorhabditis elegans (ce); Saccharomyces cerevisiae (sc); Drosophila melanogaster (dm); Homo sapiens (hs); Mus musculus (mm). Conserved residue analysis was used to make predictions regarding enzyme activity (◄).

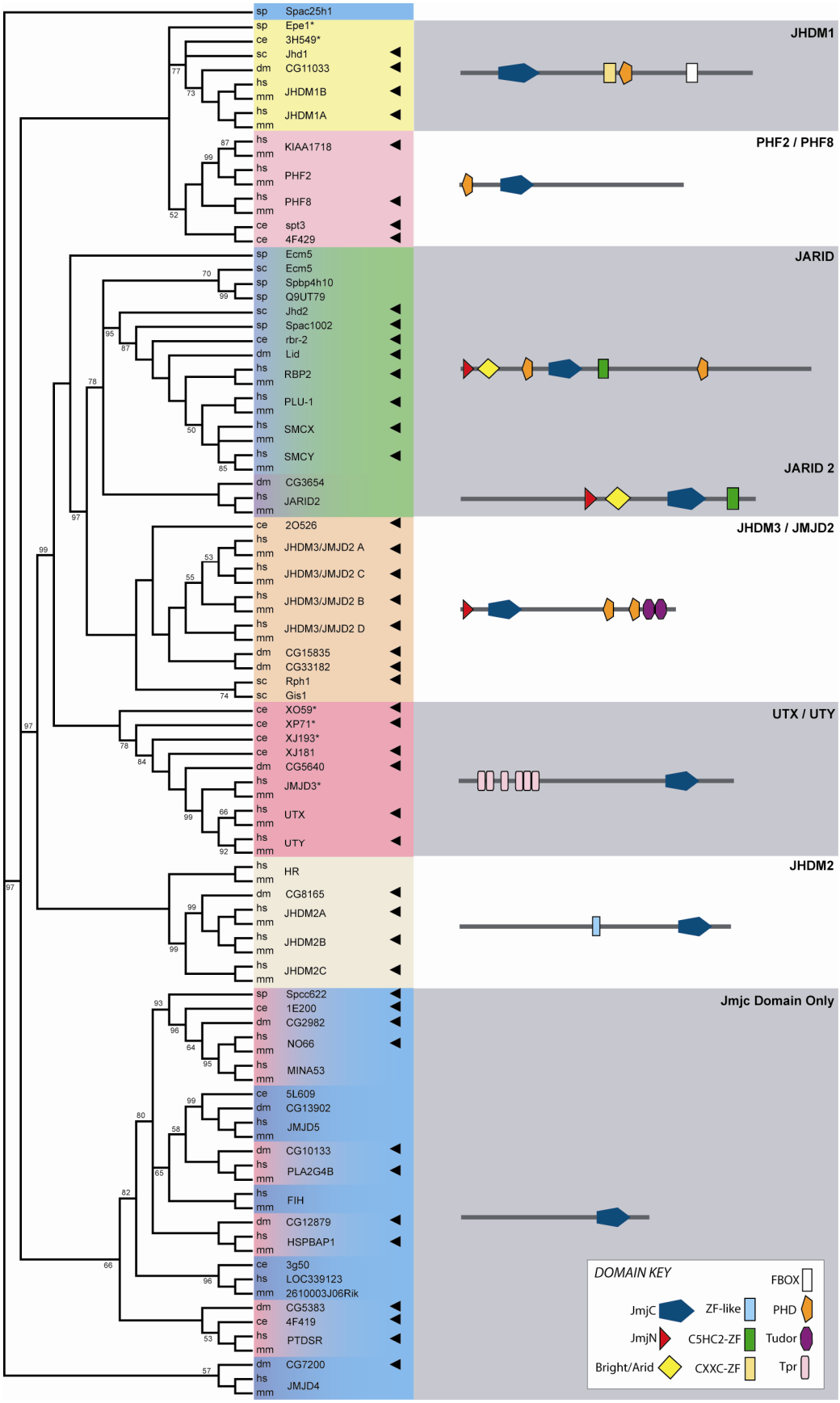
sp Spac25h1
sp Epe1*
ce 3H549*
sc Jhd1
dm CG11033
hs
mm JHDM1B
hs
mm JHDM1A
JHDM1
hs
mm KIAA1718
hs
mm PHF2
hs
mm PHF8
PHF2 / PHF8
ce spt3
ce 4F429
sp Ecm5
sc Ecm5
sp Spbp4h10
sp Q9UT79
sc Jhd2
sp Spac1002
ce rbr-2
dm Lid
hs
mm RBP2
hs
mm PLU-1
hs
mm SMCX
hs
mm SMCY
JARID
dm CG3654
hs
mm JARID2
JARID 2
ce 2O526
hs
mm JHDM3/JMJD2 A
hs
mm JHDM3/JMJD2 C
hs
mm JHDM3/JMJD2 B
hs
mm JHDM3/JMJD2 D
JHDM3 / JMJD2
dm CG15835
dm CG33182
sc Rph1
sc Gis1
ce XO59*
ce XP71*
ce XJ193*
ce XJ181
dm CG5640
hs
mm JMJD3*
hs
mm UTX
hs
mm UTY
UTX / UTY
hs
mm HR
dm CG8165
hs
mm JHDM2A
hs
mm JHDM2B
hs
mm JHDM2C
JHDM2
sp Spcc622
ce 1E200
dm CG2982
hs
mm NO66
hs
mm MINA53
ce 5L609
dm CG13902
hs
mm JMJD5
dm CG10133
hs
mm PLA2G4B
hs
mm FIH
dm CG12879
hs
mm HSPBAP1
ce 3g50
hs LOC339123
mm 2610003J06Rik
dm CG5383
ce 4F419
hs
mm PTDSR
dm CG7200
hs
mm JMJD4
Jmjc Domain Only
DOMAIN KEY
JmjC
JmjN
Bright/Arid
ZF-like
C5HC2-ZF
CXXC-ZF
FBOX
PHD
Tudor
Tpr

**Table 2-1.** Annotation of JmjC-domain-containing proteins.

| Organism | Name | Alternate Names |
|---|---|---|
| Pombe | Spac25h1 | Spac25h1.02, Jmj1 |
| Pombe | Epe1 | SPCC622.16c |
| Pombe | Spcc622 | Jmj4 |
| Pombe | Ecm5 | SPBC83.07, Lid2, jmj3 |
| Pombe | Spbp4h10.01 | SPBP19A11.06, lid2 |
| Pombe | SPAC1002 | Jmj2 |
| Pombe | Q9UT79 | SPAC343.11c, msc1 |
| Human | NO66 | C14orf169, FLJ21802 |
| Human | JHDM1B | FBXL10, CXXC2, Fbl10, PCCX2, JEMMA |
| Human | JHDM1A | FBXL11, CXXC8, DKFZP434M1735, FBL11, FBL7, FLJ00115, KIAA1004, LILIN |
| Human | JMJD5 | FLJ13798, LOC79831 |
| Human | FIH | HIF1AN, DKFZp762F1811, FIH1, FLJ20615, FLJ22027 |
| Human | HR | ALUNC, AU, HSA277165 |
| Human | HSPBAP1 | FLJ22623, FLJ39386, PASS1 |
| Human | JARID1A | RBBP2, RBP2 |
| Human | JARID1B | 2010009J12Rik, 2210016I17Rik, AW556288, D1Ertd202e, KIAA4034, PLU-1, PU |
| Human | JARID1C | Smcx, KIAA0234 |
| Human | JARID1D | Smcy |
| Human | JARID2 | Jumonji, AI317256, AU045941, C79929, C79931, Jmj |
| Human | JHDM2A | JMJD1A, DKFZp686A24246, DKFZp686P07111, JMJD1, KIAA0742, TSGA |
| Human | JHDM2B | JMJD1B, 5qNCA, C5orf7, KIAA1082 |
| Human | JHDM2C | JMJD1C, DKFZp761F0118, FLJ14374, KIAA1380, RP11-10C13.2, TRIP8 |
| Human | JHDM3A | JMJD2A, JMJD2, KIAA0677 |
| Human | JHDM3B | JMJD2B, KIAA0876 |
| Human | JHDM3C | JMJD2C, GASC1, KIAA0780, bA146B14.1 |
| Human | JHDM3D | JMJD2D, FLJ10251 |
| Human | JMJD3 | RP23-5O23.5, 1700064E03Rik, BC038313, KIAA0346, mKIAA0346 |
| Human | JMJD4 | RP23-210M6.10, 6430559I23, AU020939 |
| Human | KIAA1718 | |
| Human | LOC339123 | PP14397 |
| Human | MINA53 | DKFZp762O1912, FLJ14393, MDIG, MINA, NO52 |
| Human | PHF2 | GRC5, KIAA0662 |
| Human | PHF8 | RP13-444K19.2, DKFZp686E0868, KIAA1111, ZNF422 |
| Human | PLA2G4B | HsT16992, cPLA2-beta |
| Human | PTDSR | PSR, KIAA0585, PTDSR1 |
| Human | UTX | RP13-886N14.3, DKFZp686A03225, bA386N14.2 |

| | | |
|---|---|---|
| Human | UTY | UTY1 |
| Worm | rbr-2 | ZK593.4 |
| Worm | XJ181 | D2021.1 |
| Worm | 5L609 | C06H2.3 |
| Worm | 4F419 | psr-1, F29B9.4, tag-159 |
| Worm | 2O526 | Y48B6A.11 |
| Worm | XO59 | C29F7.6 |
| Worm | XP71 | F23D12.5 |
| Worm | CE20788 | F43G6.6, 2M97, spt-3 |
| Worm | 3H549 | T26A5.5 |
| Worm | XJ193 | tag-279, F18E9.5 |
| Worm | 4F429 | F29B9.2 |
| Worm | 1E200 | T28F2.4 |
| Worm | 3g50 | C27F2.5 |
| Fly | CG13902 | |
| Fly | CG15835 | |
| Fly | CG33182 | CG4037 |
| Fly | CG5383 | L0022859, PS, dPSR, PSR |
| Fly | CG8165 | CG31123 |
| Fly | CG5640 | |
| Fly | Lid | CG9088, lid, l(2)10424 |
| Fly | CG2982 | |
| Fly | CG12879 | |
| Fly | CG11033 | |
| Fly | CG7200 | |
| Fly | CG3654 | |
| Fly | CG10133 | |
| Cerevisiae | Jhd1 | YER051W |
| Cerevisiae | Rph1 | YER169W |
| Cerevisiae | Yjr119c | |
| Cerevisiae | Ecm5 | YMR176W |
| Cerevisiae | Gis1 | YDR096W |
| Mouse | No66 | 2410016O06Rik |
| Mouse | 2610003J06Rik | |
| Mouse | Jmjd5 | LOC77035, 3110005O21Rik |
| Mouse | Jmjd2d | Jmjd2d, 4932416A15 |
| Mouse | Kiaa1718 | BB041802, A630082K20Rik |
| Mouse | Jhdm1b | Fbxl10, Cxxc2, E430001G17, Fbl10, KIAA3014, PCCX2, mKIAA3014 |
| Mouse | Jhdm1a | Fbxl11, Cxxc8, Fbl11, Fbl7, lalina |
| Mouse | Fih | Hif1an, 2310046M24Rik, A830014H24Rik, FIH1 |
| Mouse | Hr | ALUNC, AU,N, ba,rh, rh-bmh, rhino |

| | | |
|---|---|---|
| Mouse | Hspbap1 | 3830421G21Rik |
| Mouse | Jarid1a | RBBP2, RBP2 |
| Mouse | Jarid1b | 2010009J12Rik, 2210016I17Rik, AW556288, D1Ertd202e, KIAA4034, PLU-1, PU |
| Mouse | Jarid1c | Smcx, mKIAA0234 |
| Mouse | Jarid1d | Smcy |
| Mouse | Jarid2 | AI317256, AU045941, C79929, C79931, Jmj |
| Mouse | Jhdm2a | Jmjd1a, 1700105C21Rik, C230043E16Rik, Jmjd1, TGSA, Tsga, mKIAA0742 |
| Mouse | Jhdm2b | Jmjd1b, 5830462I21Rik, mKIAA1082 |
| Mouse | Jhdm2c | Jmjd1c, 5430433L24Rik, D630035I23Rik, Jmjdic, TRIP8 |
| Mouse | Jhdm3a | Jmjd2a, JMJD2, KIAA0677 |
| Mouse | Jhdm3b | Jmjd2b, KIAA0876 |
| Mouse | Jhdm3c | Jmjd2c, GASC1, KIAA0780, bA146B14.1 |
| Mouse | Jhdm3d | Jmjd3, RP23-5O23.5, 1700064E03Rik, BC038313, KIAA0346, mKIAA0346 |
| Mouse | Jmjd4 | RP23-210M6.10, 6430559I23, AU020939 |
| Mouse | Mina53 | Mina, 1810047J07Rik, 2410057H13Rik, 3830408E23Rik, AI449204 |
| Mouse | Phf2 | |
| Mouse | Phf8 | 9830141C09Rik, mKIAA1111 |
| Mouse | Pla2g4b | HsT16992, cPLA2-beta |
| Mouse | Ptdsr | 5730436I23Rik, D11Ertd195e, PSR, PtdSerR, mKIAA0585 |
| Mouse | Utx | RP13-886N14.3, DKFZp686A03225, bA386N14.2 |
| Mouse | Uty | UTY1 |

## REFERENCES


Balciunas, D. and H. Ronne (2000). Evidence of domain swapping within the jumonji family of transcription factors. *Trends Biochem Sci* **25**(6): 274-6.

Bateman, A., L. Coin, et al. (2004). The Pfam protein families database. *Nucleic Acids Res* **32**(Database issue): D138-41.

Chenna, R., H. Sugawara, et al. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**(13): 3497-500.

Clissold, P. M. and C. P. Ponting (2001). JmjC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A2beta. *Trends Biochem Sci* **26**(1): 7-9.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**(22): 10881-90.

Cuthbert, G. L., S. Daujat, et al. (2004). Histone deimination antagonizes arginine methylation. *Cell* **118**(5): 545-53.

Dann, C. E., 3rd, R. K. Bruick, et al. (2002). Structure of factor-inhibiting hypoxia-inducible factor 1: An asparaginyl hydroxylase involved in the hypoxic response pathway. *Proc Natl Acad Sci U S A* **99**(24): 15351-6.

Dunwell, J. M. and P. J. Gane (1998). Microbial relatives of seed storage proteins: conservation of motifs in a functionally diverse superfamily of enzymes. *J Mol Evol* **46**(2): 147-54.

Elkins, J. M., K. S. Hewitson, et al. (2003). Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* **278**(3): 1802-6.

Huelsenbeck, J. P. and F. Ronquist (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**(8): 754-5.

Klose, R., K. Yamane, et al. (2006). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and 36. *Nature* **442**(7100): 312-6.

Kumar, S., K. Tamura, et al. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**(2): 150-63.

Lando, D., D. J. Peet, et al. (2002). FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev* **16**(12): 1466-71.

Lee, C., S. J. Kim, et al. (2003). Structure of human FIH-1 reveals a unique active site pocket and interaction sites for HIF-1 and von Hippel-Lindau. *J Biol Chem* **278**(9): 7558-63.

Letunic, I., R. R. Copley, et al. (2006). SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res* **34**(Database issue): D257-60.

Martin, C. and Y. Zhang (2005). The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* **6**(11): 838-49.

Metzger, E., M. Wissmann, et al. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**(7057): 436-9.

Rea, S., F. Eisenhaber, et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**(6796): 593-9.

Ronquist, F. and J. P. Huelsenbeck (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**(12): 1572-4.

Schultz, J., F. Milpetz, et al. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95**(11): 5857-64.

Shi, Y., F. Lan, et al. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**(7): 941-53.

Takeuchi, T., Y. Yamazaki, et al. (1995). Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes Dev* **9**(10): 1211-22.

Trewick, S. C., P. J. McLaughlin, et al. (2005). Methylation: lost in hydroxylation? *EMBO Rep* **6**(4): 315-20.

Tsukada, Y., J. Fang, et al. (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**(7078): 811-6.

Wang, Y., J. Wysocka, et al. (2004). Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science* **306**(5694): 279-83.

Whetstine, J. R., A. Nottke, et al. (2006). Reversal of Histone Lysine Trimethylation by the JMJD2 Family of Histone Demethylases. *Cell* **125**(3): 467-81.

Yamane, K., C. Toumazou, et al. (2006). JHDM2A, a JmjC-Containing H3K9 Demethylase, Facilitates Transcription Activation by Androgen Receptor. *Cell* **125**(3): 483-95.

# Chapter 3

## The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through $p15^{Ink4b}$

This chapter is the result of collaboration with Dr. Jin He. I directly contributed on Figure 3-1, 3-4C, 3-4D, 3-4E, S3-1, S3-2 and S3-3A. In addition I provided reagents used to generate Figures 3-4A and 3-4B.

**ABSTRACT**

The *Ink4a/Arf/Ink4b* locus plays a critical role in both cellular senescence and tumorigenesis. Jhdm1b/Kdm2b/Fbxl10, the mammalian paralogue of the histone demethylase Jhdm1a/Kdm2a/Fbxl11, has been implicated in cell cycle regulation and tumorigenesis. In this report, we demonstrate that Jhdm1b is an H3K36 demethylase. Knockdown of *Jhdm1b* in primary MEFs inhibits cell proliferation and induces cellular senescence in a pRb and p53 pathway-dependent manner. Importantly, the effect of Jhdm1b on cell proliferation and cellular senescence is mediated through de-repression of $p15^{Ink4b}$ as loss of $p15^{Ink4b}$ function rescues cell proliferation defects in Jhdm1b knockdown cells. Chromatin immunoprecipitation on ectopically expressed Jhdm1b demonstrates that Jhdm1b targets the $p15^{Ink4b}$ locus and regulates its expression in an enzymatic activity-dependent manner. Alteration of Jhdm1b level affects Ras-induced neoplastic transformation. Collectively, our results indicate that Jhdm1b is an H3K36 demethylase that regulates cell proliferation and senescence through $p15^{Ink4b}$.

## INTRODUCTION

The *Ink4a/Arf/Ink4b* locus encodes three critical cell cycle inhibitors including $p16^{Ink4a}$, $p15^{Ink4b}$ and Arf (p14 in human and p19 in mouse). The two members of the Ink4 protein family inhibit the binding of D-type cyclins to cyclin dependent kinase 4 and 6 (Cdk4/6), inhibiting the phosphorylation of retinoblastoma (Rb) family proteins and preventing G1/S phase transition in cells. The Arf protein, which shares two common $p16^{Ink4a}$ exons but contains a distinct open reading frame, is able to activate the p53 pathway through sequestration of the p53 negative regulator Mdm2 (reviewed in (Ortega, Prieto et al. 2003; Gil and Peters 2006)).

The *Ink4a/Arf/Ink4b* locus plays a critical role in determining cellular response to oncogenic signals. In normal cells, inappropriate oncogenic stimulation activates this locus and leads to cellular senescence. However, dysregulation of this locus can facilitate tumorigenesis through multiple oncogenic signaling pathways. The importance of this locus in cellular defense against tumorigenesis is further supported by evidence that *Ink4a/Arf/Ink4b* is frequently deleted or mutated in a variety of human primary tumors (Kamb, Gruis et al. 1994; Ogawa, Hirano et al. 1994; Okuda, Shurtleff et al. 1995; Kannengiesser, Dalle et al. 2007). In addition, combined deletion of *Ink4a/Arf* with *Ink4b* in mice results in a broader spectrum of tumors compared to mice with individual genetic deletions indicating that genes in this locus work synergistically to prevent tumor development and that $p15^{Ink4b}$ is a critical tumor suppressor in the absence of $p16^{Ink4a}$ (Krimpenfort, Ijpenberg et al. 2007). Since this locus controls both cellular senescence and tumorigenesis, tight regulation is crucial under physiological conditions. Although many oncoproteins and Polycomb group proteins have been shown to regulate *Ink4a/Arf* expression

(Bracken, Kleine-Kohlbrecher et al. 2007; Kotake, Cao et al. 2007), the mechanism that controls *p15*$^{Ink4b}$ expression remains unclear.

The JmjC-domain containing histone demethylase 1b (Jhdm1b) is a paralogue of the first identified histone lysine demethylase, Jhdm1a, which targets H3K36 for demethylation (Tsukada, Fang et al. 2006). This fact, as well as their high homology within the catalytic JmjC domain (79%), led us to predict that the demethylase activity against the H3K36 methyl group is conserved between both paralogues (Fig. S3-1). However, in a recent report Jhdm1b was implicated in the demethylation of H3K4me3 *in vivo* (Frescas, Guardavaccaro et al. 2007). In addition, the reported biological functions of this protein are also controversial. Although two groups have identified *Jhdm1b* as a hotspot for proviral insertion in murine tumors generated by random MMLV mutagenesis, the locus has paradoxically been identified as an oncogene and a tumor suppressor depending on the screen and functional studies used(Suzuki, Minehata et al. 2006; Pfau, Tzatsos et al. 2008). In addition, subsequent studies have reported that Jhdm1b was a negative regulator of *c-Jun* (Koyama-Nasu, David et al. 2007) or rRNA genes (Frescas, Guardavaccaro et al. 2007), further implicating Jhdm1b in tumor suppression.

In an effort to resolve these apparent discrepancies, we set out to use methods well established in our previous studies to characterize the biochemical properties as well as biological function of Jhdm1b using primary MEF cells, which maintain normal cell cycle regulatory pathways, to study protein function. We report here that Jhdm1b is an H3K36-specific histone demethylase that functions to promote cellular proliferation and inhibit cellular senescence through the silencing of the *p15*$^{Ink4b}$ tumor suppressor gene.

## RESULTS

*Jhdm1b is an H3K36me2-specific demethylase*

In an effort to explain the difference between activity prediction based on domain homology and reported *in vivo* substrate specificity (Frescas, Guardavaccaro et al. 2007), we first investigated the *in vitro* and *in vivo* catalytic activity of Jhdm1b. To this end, recombinant protein was flag-affinity purified from baculovirus-infected Sf9 cells (Fig. S3-2A), and subjected to an histone demethylase assay by measuring radioactive formaldehyde release (Tsukada, Fang et al. 2006). Of all the substrates tested, only the H3K36-specific SET2-labeled histone substrate could be demethylated by Jhdm1b, indicating that, as predicted, Jhdm1b has a similar substrate specificity to Jhdm1a (Tsukada, Fang et al. 2006) (Fig. 3-1A).

Because Jhdm1b was implicated in the demethylation of H3K4me3 *in vivo* (Frescas, Guardavaccaro et al. 2007), we tested the capacity of Jhdm1b to demethylate H3K4me3 *in vitro*. To this end, we purified a new batch of wild-type Jhdm1b along with a mutant Jhdm1b (H211A) predicted to kill the enzymatic activity (Fig. S3-2B). As controls, we also purified the H3K36me2-specific demethylase, Jhdm1a (Tsukada, Fang et al. 2006), and the H3K4me3-specific demethylase, Rbp2 (Klose, Yan et al. 2007). Incubation of these proteins with SET2-labeled histone substrate confirmed efficient formaldehyde release for Jhdm1a and Jhdm1b, but not Jhdm1b (H211A) (Fig. 3-1B). In addition, Jhdm1b exhibited about one third the activity of JHDM1A when equivalent amount of proteins were assayed. Furthermore, incubation of these proteins with substrate labeled by a mutated form of the SET7, SET7 (Y245A), which generates K4me2 and K4me3 (Xiao, Jing et al. 2003), did not show efficient release of formaldehyde above background. However, incubation of RBP2

with the same substrate resulted in formaldehyde release (Fig. 3-1B), demonstrating that Jhdm1b has no H3K4me3 demethylase activity under the assay conditions.

To further define the substrate specificity of Jhdm1b, we incubated the protein with core histones purified from HeLa cells and analyzed the product by Western blot using methylation state-specific antibodies. These experiments demonstrate that Jhdm1b, like its paralogue, JHDM1A, can specifically demethylate H3K36me2 and H3K36me1 histone substrates (Fig. 3-1C). However, it does not alter H3K4 methylation levels (Fig. 3-1C).

To characterize the substrate specificity of Jhdm1b *in vivo*, we utilized HEK293 cells selected for adherence (AD293) and overexpressed Jhdm1b and Jhdm1b (H211A) via retroviral infection. Quantitative reverse transcriptase PCR (qRT-PCR) verified similar levels of stable expression of wild-type and the mutant Jhdm1b (Fig. 3-1D). Western blot analysis demonstrated that overexpression of wild-type, but not the mutant, Jhdm1b resulted in marked decrease of H3K36me2 levels (Fig. 3-1E). Although a small decrease in H3K36me3, as well as H3K4me3 was observed, the minor decrease did not rely on a functional JmjC domain as it can also be seen in the cells over expressing the catalytic mutant. Given that overexpression of Jhdm1b in HeLa cells was previously reported to result in the decrease of H3K4me3 levels (Frescas, Guardavaccaro et al. 2007), we overexpressed Jhdm1b in HeLa cells by lentiviral infection. Western blot analysis of the histones purified from the infected HeLa cells confirmed our observation in the HEK293 cells (Fig. S3-3A). Taken together, our results indicate that Jhdm1b is an H3K36 rather than H3K4-specific demethylase.

*Jhdm1b knockdown inhibits cell proliferation and induces senescence*

We next explored the function of Jhdm1b in primary cells. Given that previous analyses of Jhdm1b tumor suppressor function had been carried out in transformed or

immortalized cell lines, we postulated that the paradoxical conclusions regarding the role of Jhdm1b in cancer might be due to the many genetic alterations necessary for the establishment of these cell lines. We therefore designed an shRNA that can target all three murine isoforms of *Jhdm1b* and introduced it into primary mouse embryonic fibroblast (MEF) cells by lentiviral infection (Jhdm1b KD). After selection, the knockdown efficiency (90%) was verified by qRT-PCR (Fig. 3-2A). Interestingly, stable knockdown of Jhdm1b resulted in a substantial decrease in cell proliferation (Fig. 3-2B), a phenotype reminiscent of that displayed by MEF cells with reduced levels of the Polycomb group protein Ring1b (Voncken, Roelen et al. 2003; Cales, Roman-Trufero et al. 2008). On the basis of this observation, shRNA directed against *Ring1b* was introduced into MEF cells (Ring1b KD), and knockdown efficiency was determined by RT-qPCR to be 90% (Fig. 3-2A). Furthermore, knockdown of *Jhdm1b* had no substantial effect on *Ring1b* levels and vice versa (Fig. 3-2A). Comparison of the cell proliferation levels of *Ring1b* KD, *Jhdm1b* KD, and control cells confirmed Ring1b's drastic effects on cell proliferation and indicated that *Jhdm1b* knockdown, while drastic, did not achieve the same level of cell growth inhibition as that of *Ring1b* knockdown (Fig. 3-2B). BrdU pulse labeling followed by flow cytometry analysis revealed that *Jhdm1b* KD resulted in a 2-fold reduction in the number of cells in S-phase as compared to control MEFs, while a 4-fold reduction is observed in the *Ring1b* KD cells (Fig. 3-2C). In addition, Ring1b has been demonstrated to play an important role in regulating cellular senescence as a component of the Polycomb repressive complex 1 (PRC1) which enacts negative regulation of the $p16^{Ink4a}$ locus (Jacobs, Kieboom et al. 1999; Cales, Roman-Trufero et al. 2008). To investigate whether *Jhdm1b* KD also contributes to pre-mature cellular senescence, *Ring1b* and *Jhdm1b* KD MEFs were analyzed for the presence of

senescence-associated β-galactosidase (SA-β-galactosidase) at different passages. While only very few cells stained positive for SA-β-galactosidase in control MEFs at passage 5, *Jhdm1b* KD MEFs underwent cellular senescence at a markedly increased rate (Fig. 3-2D) while *Ring1b* knockdown had the most drastic ability to induce senescence. Taken together, these results suggest that Jhdm1b is involved in the positive regulation of cell cycle and negative regulation of passage-dependant cellular senescence.

*Jhdm1b regulates cell proliferation through p15$^{Ink4b}$*

Cellular proliferation and senescence is tightly regulated through the p53 and pRb pathways, and both of these cellular pathways can be inhibited by the introduction of SV40 T antigen into primary cells (Goodrich, Wang et al. 1991). To determine whether the function of Jhdm1b in cell proliferation and senescence is pRb and p53 dependent, we inactivated both pathways by retroviral expression of SV40 large T antigen in MEFs and then subjected the cells to *Jhdm1b* knockdown. Knockdown of Jhdm1b does not alter the cellular proliferation of MEF cells when the p53 and pRB pathways are blocked by the SV40 large T antigen-transduced cells (Fig. S3-4A). Consistent with this result, the percentage of S-phase cells, as indicated by BrdU incorporation, is not altered by *Jhdm1b* knockdown under these conditions (Fig. S3-4B). In addition, *Jhdm1b* knockdown induced cellular senescence, as assessed by the appearance of SA-β-galactosidase staining, was also blocked in SV40 large T antigen-transduced cells (Fig. S3-4C). Collectively, these data suggest that Jhdm1b's pro-growth and anti-senescence properties act upstream of these two pathways.

Previous studies have identified several genes whose expression is linked to Rb-regulated cellular senescence and p53-dependant apoptosis (Sherr and Roberts 1999). These genes include members of the *Ink4a/Arf/Ink4b* tumor suppressor locus, *p18* (also known as

*Ink4c*), and the cell cycle regulators *p27* (also known as *Cdkn1b*) and *p21*. To determine whether loss of Jhdm1b function affected expression of any of these genes, we performed qRT-PCR in control, *Jhdm1b* KD, and *Ring1b* KD cells. Jhdm1b KD resulted in a marked upregulation of $p15^{Ink4b}$ (Fig. 3-3A). In agreement with the crucial role of Ring1b in the PRC1 complex, *Ring1b* KD resulted in a substantial upregulation of $p16^{Ink4a}$ (Jacobs, Kieboom et al. 1999; Cales, Roman-Trufero et al. 2008). In addition, we also observed a marked increase in $p15^{Ink4b}$ mRNA levels. The effect of *Jhdm1b* and *Ring1b* KD on $p15^{Ink4b}$ and $p16^{Ink4a}$ levels was also confirmed by Western blot (Fig. 3-3B). These data are consistent with the observation that the $p16^{Ink4a}$ protein has a more potent inhibitory effect on cell proliferation, as well as the notion that $p15^{Ink4b}$ functions to back-up the function of $p16^{Ink4a}$ (Krimpenfort, Ijpenberg et al. 2007). Importantly, similar to Ring1b KD, the expression of *p21*, a p53 pathway target gene, was not altered by Jhdm1b KD (Fig. 3-3A). To analyze whether $p15^{Ink4b}$ is a key mediator of Jhdm1b's function in cellular proliferation, we derived primary MEFs from $p15^{Ink4b}$ null mice (Rosu-Myles, Taylor et al. 2007) and performed knockdown of *Jhdm1b* or *Ring1b*. Loss of $p15^{Ink4b}$ function could largely rescue the slow proliferation and low BrdU incorporation caused by *Jhdm1b* knockdown (Fig. 3-3C,D). Thus, we conclude that Jhdm1b regulates cellular proliferation and senescence by negatively regulating the expression of the $p15^{Ink4b}$ tumor suppressor gene in primary cells.

*Jhdm1b regulates $p15^{Ink4b}$ expression through H3K36 demethylation*

H3K36 methylation, which has been linked to active gene transcription in organisms from yeast to humans, is present within the coding regions of genes being actively transcribed, and tends to peak towards the 3' end of transcribed regions (Bannister, Schneider et al. 2005; Martin and Zhang 2005; Morris, Shibata et al. 2005; Rao, Shibata et al. 2005).

The mechanism by which high levels of H3K36me2 are excluded from the promoter of active genes, as well as the functional significance of this observation is still not clear in higher eukaryotes. However, it has been demonstrated that SET2 association with elongating RNAPII is at least partly responsible for this phenomenon in yeast (Morris, Shibata et al. 2005). To investigate whether the demethylase activity of Jhdm1b directly contributes to *p15*$^{Ink4b}$ regulation and cellular proliferation, we attempted to rescue the *Jhdm1b* KD MEFs with siRNA resistant wild-type F-Jhdm1b as well as catalytically deficient mutant, F-Jhdm1b (H211A). After confirming equal expression of the rescue constructs (Fig. S3-5), we analyzed their effects on cellular proliferation. Re-introduction of wild-type Jhdm1b rescued the growth defects, whereas re-introduction of the catalytically defective mutant did not (Fig. 3-4A). In addition, re-introduction of wild-type Jhdm1b, but not the catalytic mutant, restored *p15*$^{Ink4b}$ expression to control levels (Fig. 3-4B). Thus, both normal cell proliferation and *p15*$^{Ink4b}$ repression depends on the demethylase activity of Jhdm1b.

Having established that transcriptional repression of *p15*$^{Ink4b}$ depends on the H3K36-demethylase activity of Jhdm1b, we used chromatin immunoprecipitation (ChIP) assays to decipher whether *p15*$^{Ink4b}$ is a direct target of Jhdm1b. Because none of the available Jhdm1b antibodies worked in immunoprecipitation in our hands, we resolved to retrovirally express F-Jhdm1b in primary MEF cells and performed ChIP assays using Flag antibodies. F-Jhdm1b was found to be specifically enriched 3-fold above background in regions just upstream (amplicon 1) and surrounding the transcription start site (amplicon 2) of *p15*$^{Ink4b}$ when compared with mock infected cells (Fig. 3-4C,D). However, the protein was not enriched within the *p15*$^{Ink4b}$ intron (amplicon 3) (Fig. 3-4C,D). ChIP analysis of H3K36me2 levels at these same regions in *Jhdm1b* KD and control cells revealed that KD of *Jhdm1b* resulted in

an increase in the H3K36me2 levels across the locus when compared to control cells (Fig. 3-4E), suggesting that Jhdm1b might also play an active role in the demethylation of H3K36me2 downstream of the transcription start site. However, it has not escaped our attention that increased transcription of the gene may itself contribute to increases in H3K36me2 within the coding region. Furthermore, rescue of the KD by wild-type Jhdm1b resulted in H3K36me2 levels below those of control cells (Fig. 3-4E). In addition, H3K36me2 levels within the *Gapdh* locus were unchanged for all samples (data not shown). Taken together, these data suggest that $p15^{Ink4b}$ is a direct Jhdm1b target and that Jhdm1b regulates $p15^{Ink4b}$ expression through active demethylation of H3K36.

*Jhdm1b cooperates with Ras to induce oncogenic transformation*

All of the data presented above indicates that Jhdm1b functions as a proto-oncogene through its ability to repress transcription of the $p15^{Ink4b}$ locus. In an effort to further establish the oncogenic potential of Jhdm1b, we assessed the contribution of the protein to colony formation using knockdown or overexpression in *p53* null MEFs followed by superinfection of retroviral H-Ras12V virus. Results of soft-agar colony formation analysis indicate that co-expression of RAS and Jhdm1b in *p53* null MEF cells results in an increase in Ras induced oncogenic transformation (Fig. 3-5A). Conversely, knockdown of *Jhdm1b* inhibited Ras-induced colony formation (Fig. 3-5B). Together these data support the notion that *Jhdm1b* acts as a proto-oncogene in primary fibroblast cells.

## DISCUSSION

In this report, we demonstrate that Jhdm1b, like its paralogue Jhdm1a, functions as an H3K36 demethylase *in vitro* and *in vivo*. Recombinant Jhdm1b purified from baculovirus infected Sf9 cells demethylates H3K36me2 *in vitro* in a radioactive formaldehyde release assay and in histone Western blot analysis (Fig. 3-1) and Western blotting and ChIP analysis indicate that overexpression of Jhdm1b results in global H3K36me2 demethylation (Fig. 3-1E), as well as gene-specific H3K36me2 demethylation (Fig. 3-4) in different cell types including HeLa and HEK293 cells (Fig. S3-3A). In addition, the previously suggested nucleolar staining pattern of Jhdm1b could not be observed in either HeLa cells (Frescas, Guardavaccaro et al. 2007) (Fig. S3-3B), or AD293 cells (Fig. S3-5B).

In addition to the substrate specificity of Jhdm1b, its role in tumorigenesis has also been a point of contention. One of the earliest reports of Jhdm1b function came out of a genetic screen for tumor suppressor genes in mouse lymphomas (Suzuki, Minehata et al. 2006). The authors identified several bi-allelic retroviral insertion events at the *Jhdm1b* locus when several of the induced lymphomas were analyzed. However, further analysis of additional tumor samples revealed locus insertion that left the *Jhdm1b* coding region intact. This leaves open the possibility that the tumor manifesting retroviral insertions could either activate or suppresses *Jhdm1b* function. Therefore, the potential of *Jhdm1b* to serve as an oncogene or tumor suppressor was not resolved in this study. In addition some indirect evidence suggested Jhdm1b may act as a tumor suppressor, including a link to the negative regulation of *c-Jun* (Koyama-Nasu, David et al. 2007), as well as a description of its role in the negative regulation of rRNA genes (Frescas, Guardavaccaro et al. 2007). However, these experiments were carried out in various cell lines in which critical pathways important for

cellular senescence and tumorigenesis (such as those regulated by *Ink4a/Arf/Ink4b*) are also disrupted. Thus further evidence would be required to support the relevance of previously identified gene targets in a normal context. A recent report that screened 44 random MMLV induced T cell lymphomas identified the *Jhdm1b* locus as an insertion hotspot (Pfau, Tzatsos et al. 2008). This study revealed multiple, orientated provirus insertions upstream of the canonical *Jhdm1b* promoter and further implicated a role for Jhdm1b overexpression in the immortalization of primary MEF cells.

In this study, we provide several lines of evidence that support *Jhdm1b* may indeed be a proto-oncogene. First, we demonstrate in primary MEFs that knockdown of Jhdm1b resulted in cell proliferation defects and increased senescence (Fig. 3-2). Furthermore, we demonstrate that Jhdm1b contributes to the regulation of cell proliferation and senescence by directly repressing the expression of the $p15^{Ink4b}$ tumor suppressor (Figs. 3-3 & 3-4). Importantly, $p15^{Ink4b}$ appears to be a major target that mediates Jhdm1b function in cellular proliferation since loss of $p15^{Ink4b}$ function can largely rescue the *Jhdm1b* knockdown effects (Fig. 3-4). Finally, Jhdm1b can cooperate with Ras to transform primary MEF cells (Fig. 3-5). Therefore, we conclude that Jhdm1b is indeed a proto-oncogene which functions at least partially by controlling the expression of the $p15^{Ink4b}$ tumor suppressor through removal of H3K36me2.

## EXPERIMENTAL PROCEDURES

*Cloning, cell line generation:* All of the cell lines used in this study were maintained in Dulbeccos's modified Eagle's medium supplemented with 10% (v/v) fetal bovine serum and 1% (v/v) penicillin/streptomycin. Primary MEFs harvested from day 13.5 C57BL/6 and *p15*$^{Ink4b}$ null mouse embryos were plated into a P75 flask. The confluent cells were frozen down and considered passage 1. The cells were split at 1:5 for each passage. Cell growth was measured by plating 5x10$^5$ MEFs on a 100mm plate in triplicate for each group. The cell number was counted using a hemocytometer at the time points indicated. Selection and stable maintenance of exogenous plasmids was accomplished in the presence of either 1μg ml$^{-1}$ (AD293 cells) or 4μg ml$^{-1}$ (MEF cells) puromycin. PCR amplification of an EST clone corresponding to full length Jhdm1B (NM_001003953) was cloned into a modified pFastBacHT B vector (Invitrogen) containing an N terminal flag epitope tag for baculovirus generation as per manufacturer's protocol (Invitrogen). Full length Jhdm1B was subcloned into pMSCVpuro vector (Clontech) for stable cell line generation using the manufacturer's protocol. Jhdm1b (H211A) mutant was generated by employing overlapping PCR/subcloning. Stable knockdown was achieved using a lentiviral system obtained from the NIH AIDS Research and Reference Reagent Program. The mouse U6 promoter was cloned from mouse genomic DNA and inserted into the NotI site of pTY-EF1a-nLacZ. For the LV-U6 shRNA-Pgk-Pac construct, the Pgk-Pac cassette at NotI/EcoRI sites replaced the EF1a-nLacZ cassette. The hairpin RNA targeting Jhdm1b (5'-GCTCCAACTCAGTTACTGT-3'), Ring1b (5'-GCAGTACACCATTTACATA-3') and control (5'-GTTCAGATGTGCGGCGAGT-3') were cloned into BBSI/HindIII sties under the U6 promoter (Cao, Wang et al. 2008). To generate wild type and mutant Jhdm1b (H211A)

rescue constructs, the siRNA target site of Jhdm1b was mutated to ATTGCAGTTGAGTTACTGT by PCR mutagenesis. The siRNA resistant wild type and mutant Jhdm1b cDNAs were PCR amplified and cloned into either SpeI/EcoRI site of LV vector or NotI/XbaI of RV vector.

*In vitro histone demethylase formaldehyde release assay:* Histone substrates were radiolabeled and formaldehyde release assays were performed as previously described(Zhang, Tamaru et al. 2002; Tsukada, Fang et al. 2006). Unless otherwise stated, 5μg of Flag purified recombinant protein was incubated in the presence of labeled substrate corresponding to 60,000 input counts and demethylase buffer (50 mM HEPES-KOH (pH 8.0), 70 mM $Fe(NH4)_2(SO4)_2$, 1 mM α-ketoglutarate, 2 mM ascorbate) for one hour at 37˚C. Released, labeled formaldehyde was extracted using a modified NASH technique and subjected to scintillation counting. Data is presented as counts per minute in the extracted sample.

*Western blot analysis and immunostaining:* Total protein was extracted by RIPA buffer. Purified native histones from HeLa cells or acid extracted histones from the indicated cell lines were prepared as previously described(Wang, Wang et al. 2004). Antibodies against specific methylation states were used at dilutions ranging from 1:250 – 1:1000: H3K36me1(Abcam 9048), H3K36me2(Tsukada, Fang et al. 2006) H3K36me3(Abcam 9050), H3K4me1(Abcam 8895), H3K4me2 (Abcam 7766), H3K4me3(Abcam 8580). Blots were normalized using an antibody against pan H3 (Abcam 1791). Anti-p21(Santa Cruz SC-397), anti-$p15^{Ink4b}$ (Cell Signaling 4822), and anti-$p16^{Ink4a}$ (Santa Cruz SC-1207) were used at dilution of 1:1000 for Western Blot. Indirect immunostaining was carried out using primary MEFs and Hela cells. The cells were plated onto cover slips in 6 well plates after LV

transduction, fixed 48 hours post transduction for 20 min in 4% (w/v) paraformaldehyde, washed with three times with PBS, and subsequently permeabilized for 20 min in 0.5% (v/v) Triton-X-100/PBS. Permeablized cells were blocked in 3% (w/v) BSA/PBS for 30min and incubated with Flag monoclonal M2 antibody (Sigma) at 1:1000 dilution in a humidified chamber for 3 hours. After incubation, cells were washed 3 times and incubated with FITC or Rhodamine conjugated secondary antibodies (Jackson ImmunoResearch Laboratories) at dilution of 1:500. Cells were washed twice with PBS, stained with 4,6-diamidino-2-phenylindole dihydrochloride (DAPI), and mounted on glass slides in fluorescence mounting medium (DAKO). Slides were analyzed on an Axiovert 200 fluorescent microscope (Zeiss).

*BrdU incorporation and cycle cycle analysis:* MEFs were grown in the presence of 10 μM BrdU for 60min. The cells were harvested and fixed overnight using 75% ethanol. Fixed cells were stained with FITC conjugated mouse anti-BrdU antibody (BD biosciences, 347583) and propidium iodide. The stained cells were analyzed by flow cytometry (BD FACSCalibur) and data was analyzed using WinMDI version 2.9 (TSRI flowcytometry Software).

*Senescence associated-β-galactosidase assay:* Cells were washed twice with PBS and immersed in fixation buffer (2% (w/v) formaldehyde, 0.2% (w/v) glutaraldehyde in PBS) for 10 min. After two additional PBS washes the cells were allowed to stain overnight in staining solution (40 mM citric acid/sodium phosphate pH 6.0, 150 mM NaCl, 2.0 mM $MgCl_2$, 1 mg $ml^{-1}$ x-gal).

*Reverse transcription and qPCR:* RNA was extracted and purified from cell lines using Qiashredder (Qiagen) and RNeasy (Qiagen) spin columns; DNase treated (Promega RQ1 Dnase) and cleaned up using Qiagen RNeasy column (Qiagen). 1μg of RNA was subjected to reverse transcription using random primers (Promega) and Improm-II reverse transcription

kit (Promega). cDNA levels were assayed via Real time PCR using SYBR GreenER (Invitrogen) and analyzed on an ABI 7300 Real Time PCR System with SDS software version 1.3.1. qRT-PCR primer sequences are available in Table 3-1.

*ChIP assay:* ChIP assays employing flag-antibody were carried out as previously reported (Zeng, Vakoc et al. 2006) with the following modifications: 20µl of M2 agarose (Sigma) was used in the immunoprecipitation and chromatin-bound beads were washed three times each in TSEI (0.1% [w/v] SDS, 1% [v/v] Triton-X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris [pH 8.1]) TSEII (0.1% [w/v] SDS, 1% [v/v] Triton-X-100, 2mM EDTA, 500 mM NaCl, 20 mM Tris [pH 8.1]) and TSEIII (0.25M LiCl, 1% [v/v] NP-40, 1% [w/v] deoxycholate, 1mM EDTA, 10mM Tris [pH 8.1]) followed by two washes in TE. Histone modification ChIPs were carried out as previously reported(Cao, Tsukada et al. 2005). ChIP DNA was analyzed via qPCR and data are presented as percentage of input as determined using Applied Biosystems's SDS software Absolute Quantification protocol. Primer sequences are available in Table 3-1.

*Soft agar colony assay:* p53 null MEFs were transduced with different LVs. 48 hours after transduction, the cells were superinfected with retroviral H-Ras12V virus. 5000 cells were mixed in the 0.35% (w/v) top agar and plated onto 0.5% (w/v) basal agar. 14 days after plating, the cells were stained with 0.005% (w/v) crystal violet and colony number was counted.

*Standard error reporting:* All *in vitro* formaldehyde release assays and cell based assays were performed in triplicate and error bars represent the standard deviation of three independent experiments. ChIP assays and qRT-PCR experiments were repeated at least twice and data is reported for one of the biological replicates. Error bars represent the

standard deviation of three qPCR reactions as determined by SDS 1.3.1 software (Applied Biosystems). Soft agar colony assays were performed and error bars represent the standard deviation of large colony numbers between three separate plating replicates.

**Figure 3-1.** Jhdm1b is an H3K36-specific demethylase *in vitro* and *in vivo*. (**A**) *In vitro* formaldehyde release assay using histone substrates generated with the indicated methyltransferases and incubated with either recombinant Flag-Jhdm1b (empty bars) or no enzyme control (black bars). (**B**) *In vitro* formaldehyde release assay using recombinant Jhdm1b, Jhdm1b(H211A), JHDM1A and RBP2 incubated with H3K36-methylated (empty bars) or H3K4-methylated (black bars) substrates generated with SET2 or SET7(Y245A) methyltransferase, respectively. (**C**) HeLa cell core histones were incubated in the presence (+) or absence (-) of recombinant Jhdm1b. The resulting reaction was subjected to Western blotting with the indicated modification state-specific antibodies. (**D**) *Jhdm1b* mRNA levels from AD293 cells stably over-expressing F-Jhdm1b, F-Jhdm1b(H211A), or empty vector control cells. Samples were normalized relative to *Gapdh* and the *Jhdm1b* expression level in the empty vector control cells was arbitrarily set to 1. (**E**) Acid extracted histones were prepared from AD293 cells stably overexpressing F-Jhdm1b or F-Jhdm1b(H211A) and subjected to Western blotting using the indicated methyl-state specific histone antibodies. All error bars represent s.d. (*n*=3).

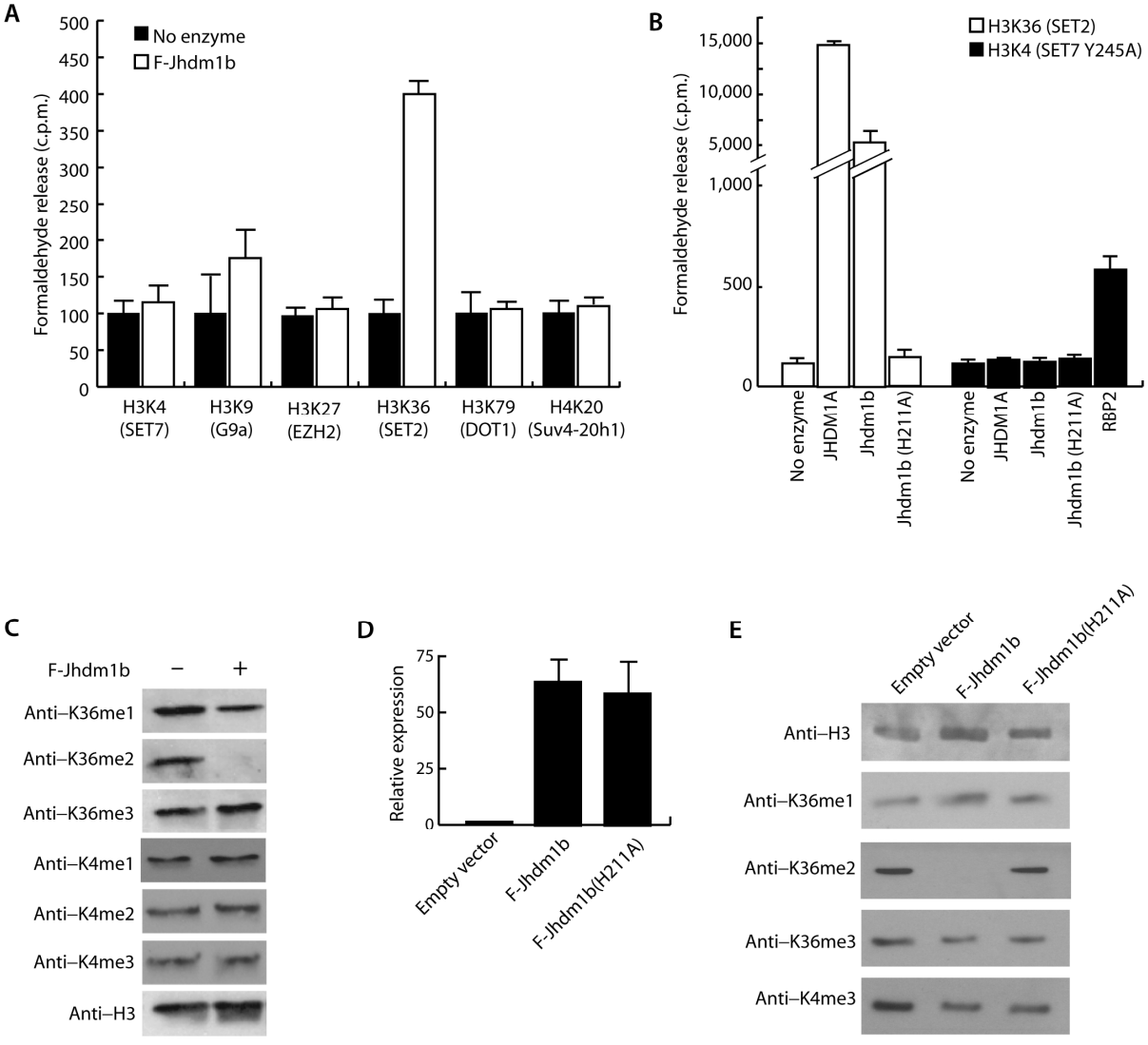

**Figure 3-2.** *Jhdm1b* knockdown in primary MEFs inhibits cell proliferation and induces cellular senescence. (**A**) Primary MEFs at passage 1 were transduced with lentiviral control (black bars), *Jhdm1b* (Jhdm1b KD, grey bars) or *Ring1b* (Ring1b KD, empty bars) shRNA and further selected by puromycin for 48 hours. mRNA levels were determined by quantitative real-time PCR (qRT-PCR). Samples were normalized against levels of *Gapdh*. The *Jhdm1b* or *Ring1b* expression level in empty vector control cells was arbitrarily set to 1. (**B**) Primary MEF cells transduced at passage 1 with lentiviral control (diamonds), *Jhdm1b* (Jhdm1b KD, squares), and *Ring1b* (Ring1b KD, triangles) shRNA and further selected by puromycin for 48 hours. $5x10^5$ cells were plated and cell number was counted at different time points using a hemocytometer. (**C**) Primary MEFs were prepared as in (B), pulse labeled with BrdU and stained with PI and FITC conjugated anti-BrdU antibody. The percentage of BrdU positive cells was determined by flow cytometry for control (black bars), Jhdm1b KD (grey bars) and Ring1b KD (empty bars). (**D**) Cells prepared as in (B) were split at 1:5 for 5 passages and senescence associated β-galactosidase activity at pH 6.0 was determined by cellular staining. Data is presented for passage 1 (P1) and passage 5 (P5) cells. All error bars represent s.d. ($n$=3).

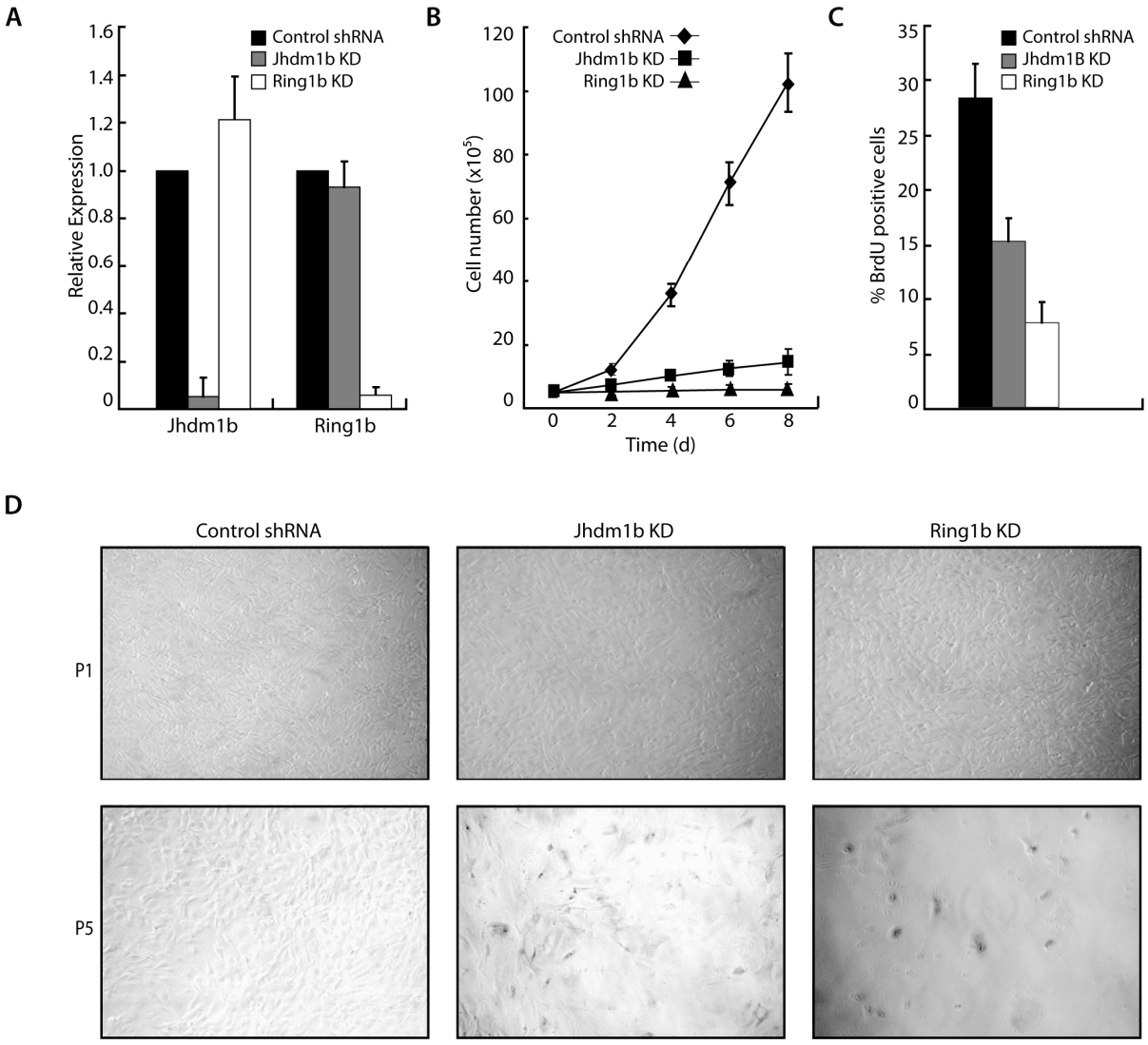

**Figure 3-3.** Jhdm1b regulates cell proliferation through $p15^{Ink4b}$. (**A**) Primary MEFs at passage 1 were transduced with lentiviral control (black bars), *Jhdm1b* (Jhdm1b KD, grey bars) and *Ring1b* (Ring1b KD, white bars) shRNA and further selected by puromycin for 48 hours. The expression level of the indicated genes was determined by qRT-PCR. Samples were normalized against levels of *Gapdh* and the expression level of each gene in empty vector control cells was arbitrarily set to 1. (**B**) Total protein was collected from cells prepared as in (A) and subjected to Western Blot analysis using antibodies against $p15^{Ink4b}$, $p16^{Ink4a}$, p21 and tublin. (**C**) Primary MEF cells isolated from $p15^{Ink4b}$ null mice at passage 1 were transduced with lentiviral control (diamonds), *Jhdm1b* (Jhdm1b KD, squares), and *Ring1b* (Ring1b KD, triangles) shRNA and further selected by puromycin for 48 hours. $5x10^5$ cells were plated and cell number was counted at different time points using a hemocytometer. (**D**) Primary MEFs were prepared as in (C), pulse labeled with BrdU and stained with PI and FITC conjugated anti-BrdU antibody. The percentage of BrdU positive cells was determined by flow cytometry for control (black bars), Jhdm1b KD (grey bars) and Ring1b KD (empty bars). All error bars represent s.d. (*n*=3).

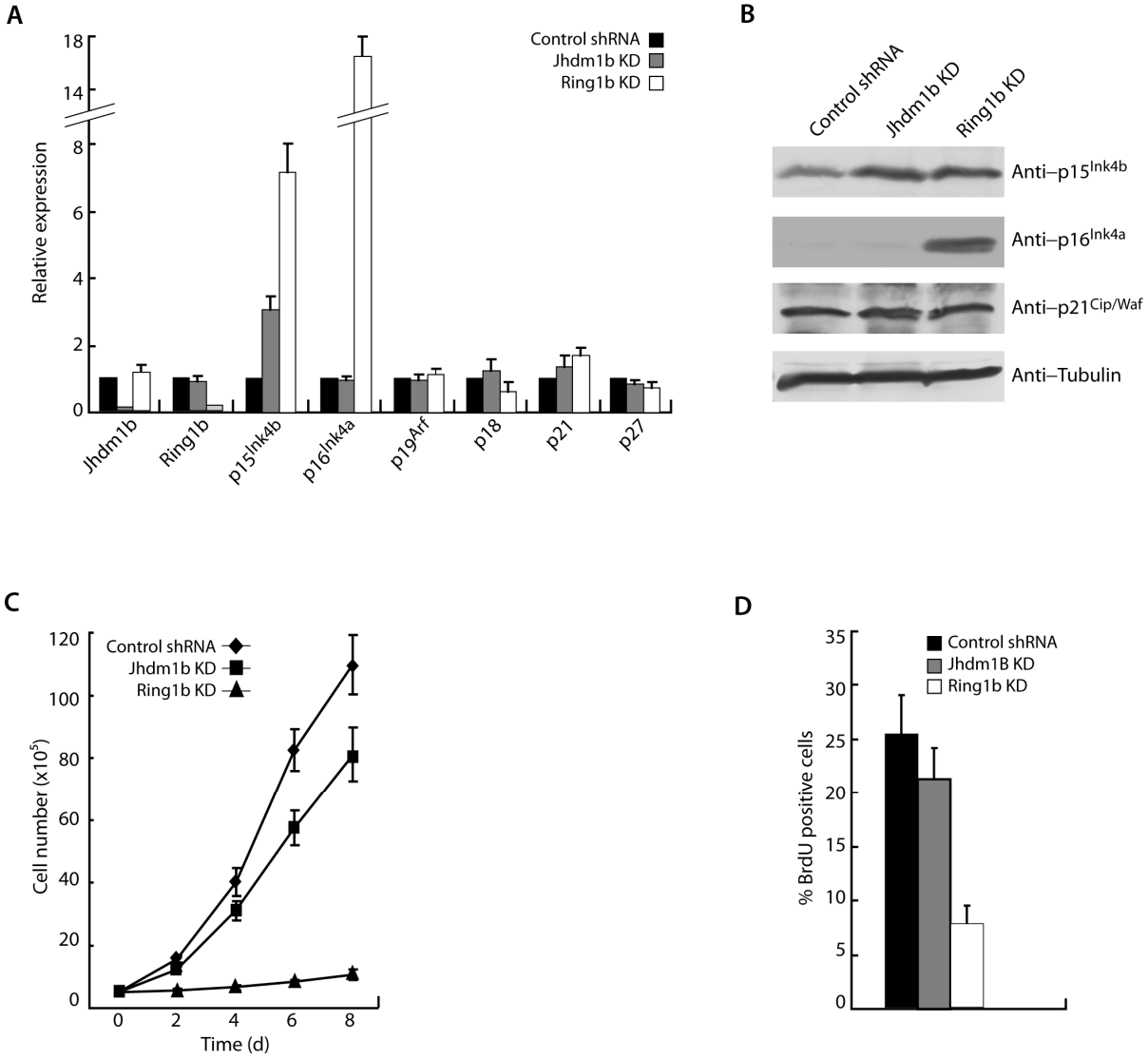

**Figure 3-4.** Jhdm1b regulates cell proliferation and *p15*$^{Ink4b}$ expression in a histone demethylase activity-dependent manner. (**A**) Primary MEF cells were transduced with lentiviral vectors carrying control shRNA (diamonds), *Jhdm1b* shRNA (Jhdm1b KD, squares), and *Jhdm1b* shRNA with expression of F-Jhdm1b (Jhdm1b KD/F-Jhdm1b, circles) or F-Jhdm1b(H211A) (triangles). $5x10^5$ cells were plated 48 hours after transduction and maintained in culture for 8 days. Cell number was counted using a hemocytometer at the indicated time points. (**B**) mRNA was collected from cells prepared as in (A) and qRT-PCR was carried out to determine *p15*$^{Ink4a}$ expression level in each sample. Samples were normalized against levels of *Gapdh* and the expression level of each gene in empty vector control cells was arbitrarily set to 1. (**C**) Schematic representation of the *p15*$^{Ink4b}$ locus in mouse indicating the genomic structure of the locus (exons demarcated by black boxes), as well as the location of the three amplicons studied in ChIP experiments. (**D**) ChIP experiment utilizing chromatin prepared from primary MEF cells transduced with retrovirus encoding F-Jhdm1b (empty bars) or control virus (black bars) were carried out using antibody against Flag. F-Jhdm1b binding was assayed via qPCR at the three genomic regions depicted in (C). (**E**) ChIP experiment utilizing chromatin from primary MEFs prepared as in (A) was carried out using antibody specific for H3K36me2. H3K36me2 levels were assayed via qPCR at the three genomic regions depicted in (C). Error bars represent s.d. (n=3).

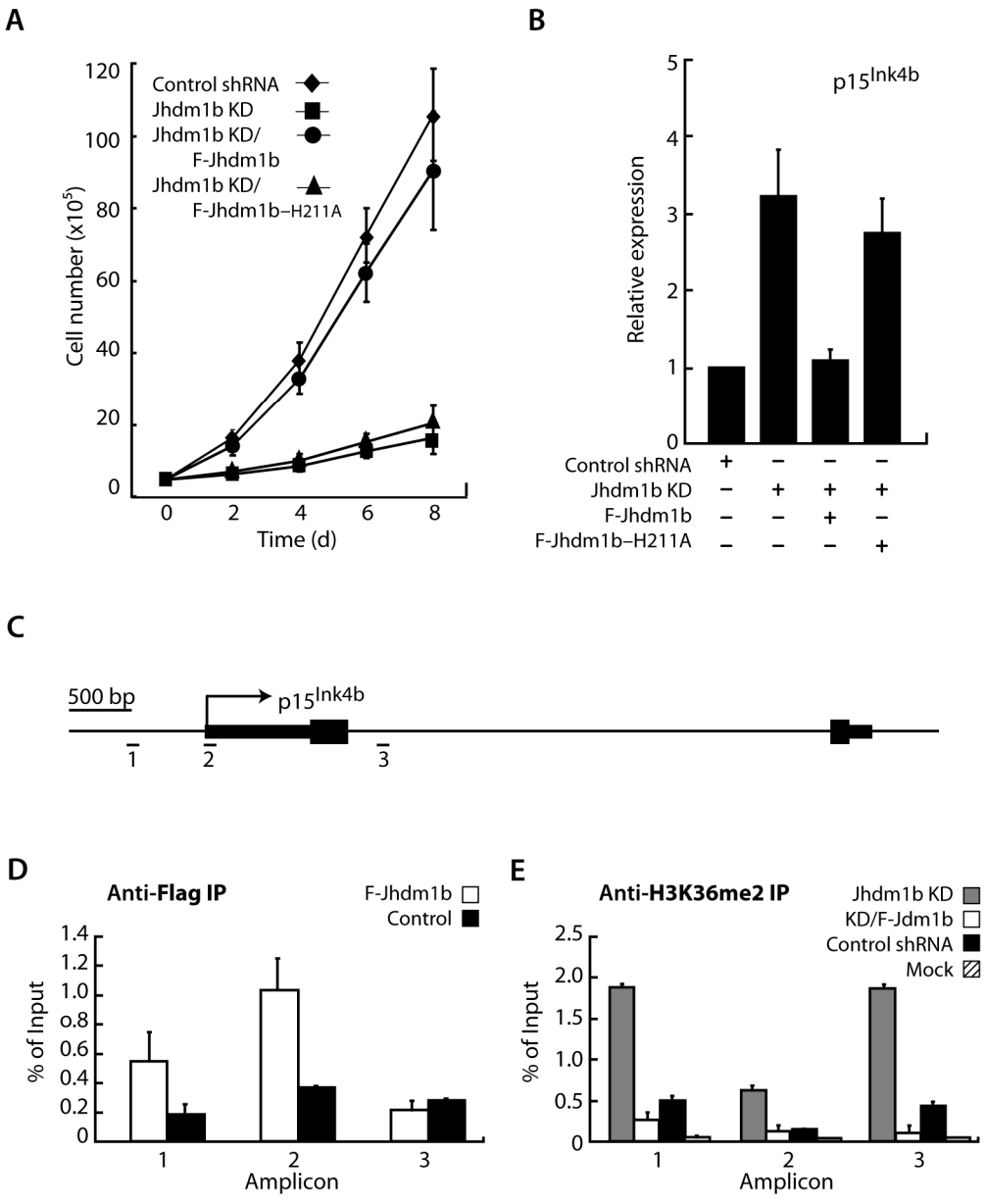
A
Control shRNA
Jhdm1b KD
Jhdm1b KD/
F-Jhdm1b
Jhdm1b KD/
F-Jhdm1b–H211A
Cell number (x10^5)
Time (d)
B
p15Ink4b
Relative expression
Control shRNA
Jhdm1b KD
F-Jhdm1b
F-Jhdm1b–H211A
C
500 bp
p15Ink4b
D
Anti-Flag IP
F-Jhdm1b
Control
% of Input
Amplicon
E
Anti-H3K36me2 IP
Jhdm1b KD
KD/F-Jdm1b
Control shRNA
Mock
% of Input
Amplicon

**Figure 3-5.** Jhdm1b facilitates Ras induced neoplastic transformation. (**A**) Primary MEF cells were transduced with mock or F-Jhdm1b expressing lentiviral vector followed by superinfection with retroviral H-Ras12V. 5000 cells were plated on soft agar and analyzed 14 days later. Graph represents quantification of large colonies on H-Ras transduced plates. (**B**) Primary MEF cells were transduced with lentiviral vectors carrying control shRNA (Control shRNA) or *Jhdm1b* shRNA (Jhdm1b KD) followed by puromycin selection and retroviral H-Ras12V transduction. 5000 cells were plated on soft agar and analyzed 14 days later. Graph represents quantification of large colonies on H-Ras transduced plates. All error bars represent s.d. (n=3).

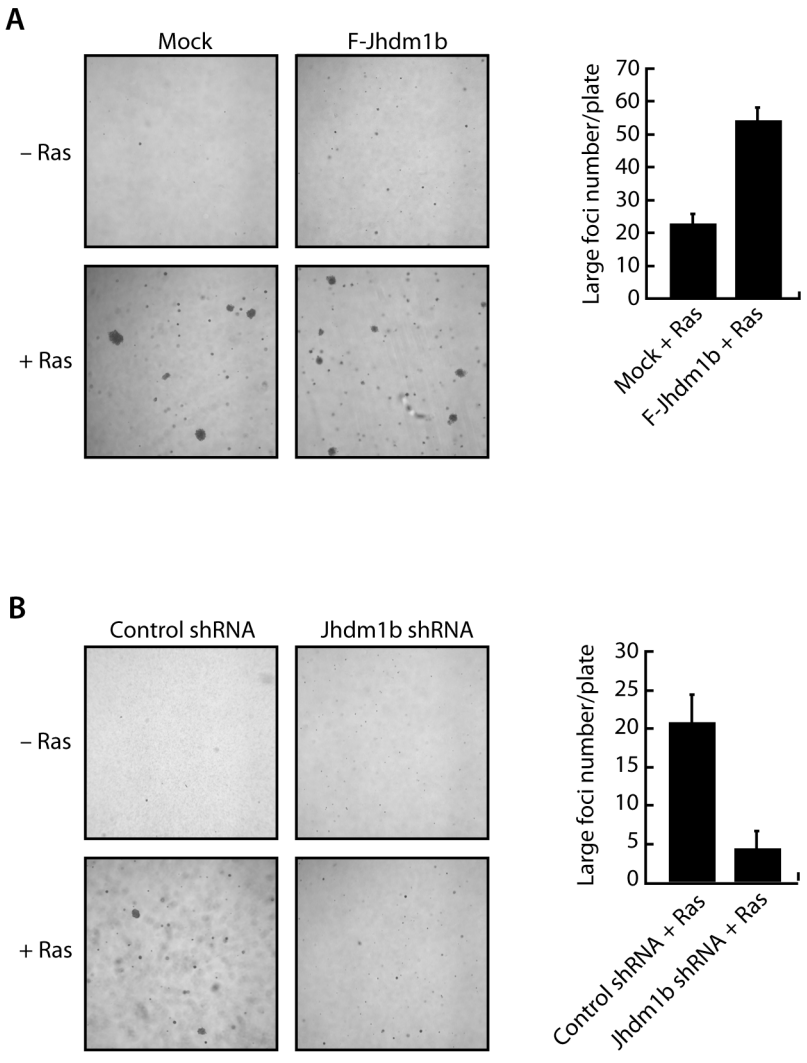

**Figure S3-1.** Amino acid alignment of the JmjC domain of Jhdm1a and Jhdm1b reveals substantial homology and conservation of residues important for demethylase activity. α-ketoglutarate and Fe(II) cofactor binding sites are indicated (blue and red, respectively). The location of the amino acid substitution resulting in the demethylase activity dead point mutant H211A is indicated in green.

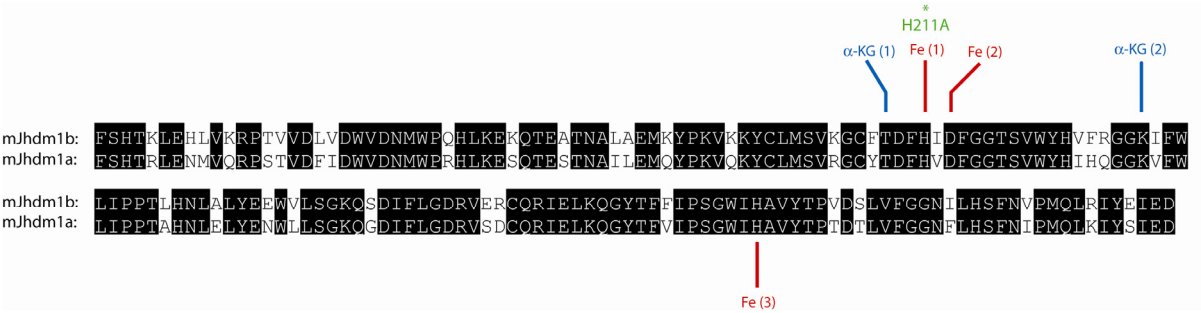

**Figure S3-2.** Coomassie staining of recombinant proteins used in the demethylase assays. Recombinant proteins purified from infected Sf9 cells was subjected to SDS PAGE separation followed by Coomassie staining. Proteins used in Fig. 3.1A (**A**) and Fig. 3.1 B (**B**) are shown.

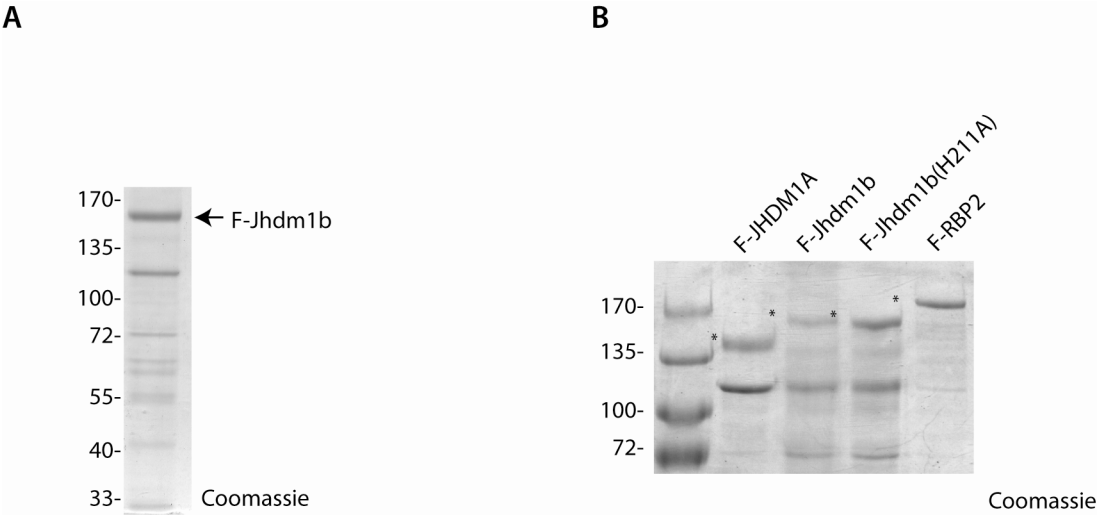

**Figure S3-3.** Overexpression of Jhdm1b in HeLa cells results in a global decrease of H3K36me2, but not H3K4me3, levels. (**A**) Acid extracted histones and whole cell lysate were prepared from puromycin selected HeLa cells infected twice with lentivirus carrying F-Jhdm1b. Histone extracts were subjected to Western blotting utilizing antibodies against histone H3, H3K36me2, and H3K4me3. Whole cell extract was probed for the presence of exogenous Jhdm1b using flag antibody. (**B**) Immunostaining of the exogenous F-Jhdm1b in transduced HeLa cells revealed a uniform nuclear staining pattern.

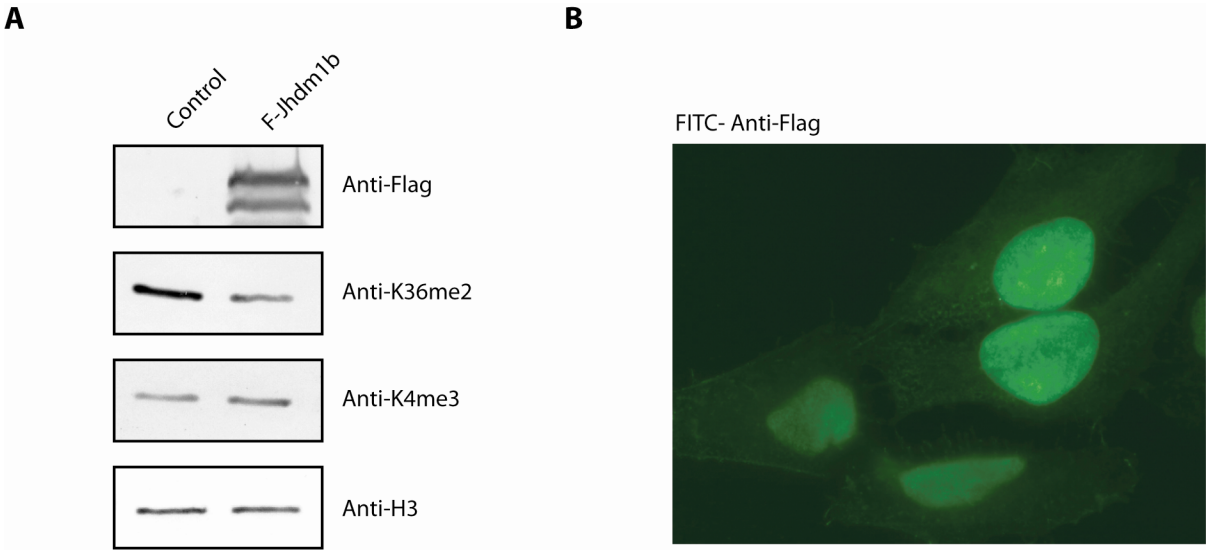

**Figure S3-4.** Overexpression of SV40 large T antigen abrogates the cell proliferation, cell cycle and cellular senescence phenotype associated with *Jhdm1b* KD. (**A**) 5x105 primary MEF cells were plated after transduction at passage 1 with lentiviral SV40 large T antigen followed by superinfection with either lentiviral control (black diamonds) or *Jhdm1b* shRNA (Jhdm1b KD, red squares). Cells were allowed to proliferate for eight days and cell number was counted at the indicated time points using a hemocytometer. (**B**) Primary MEF cells were prepared as in (a), pulse labeled with BrdU, and stained with PI and FITC conjugated anti-BrdU antibody. The percentage of BrdU positive cells was determined by flow cytometry. (**c**) Cells were prepared as in (a) and maintained in culture for 8 passages (split 1:5). The intracellular senescence associated β-galactosidase activity at pH 6.0 was stained at each passage and data is shown for passage 1 and passage 8 cells.

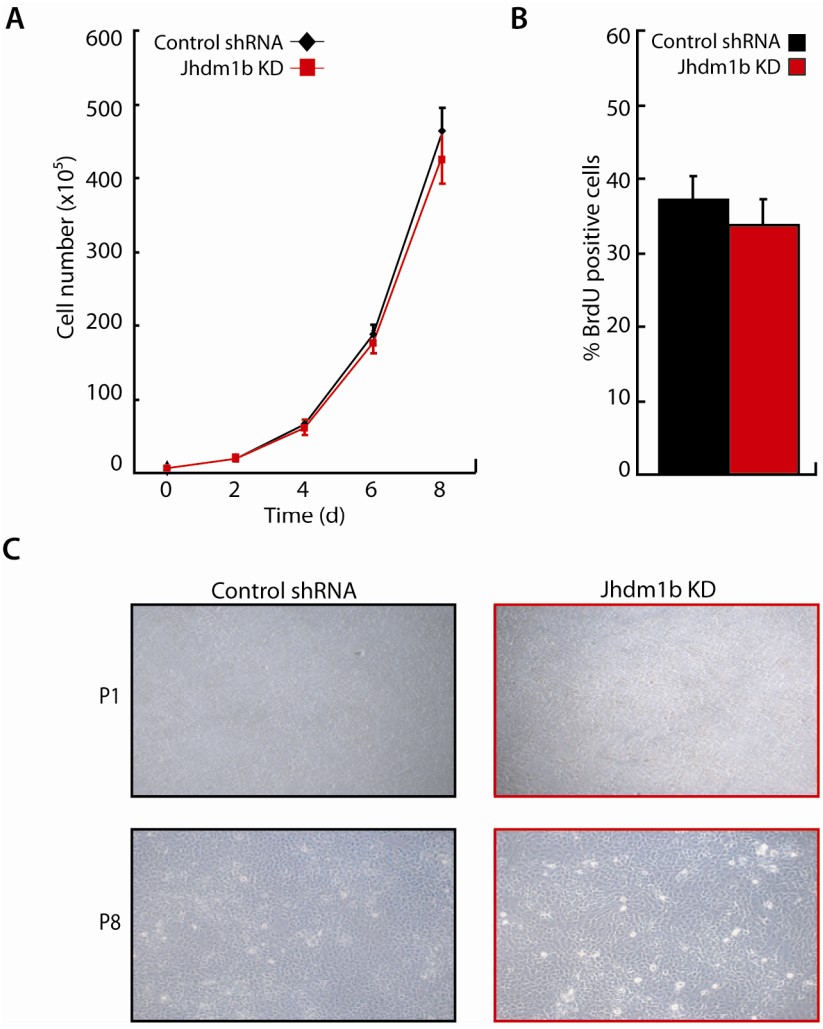

**Figure S3-5.** Jhdm1b wild-type and mutant rescue constructs express similar levels of flag-tagged protein. (**A**) Western blot analysis of exogenous F-Jhdm1b and mutant F-Jhdm1b (H211A) expression in *Jhdm1b* KD MEF cells. (**B**) Immunostaining of the exogenous F-Jhdm1b and mutant F-Jhdm1b (H211A) expression in *Jhdm1b* KD MEF cells.

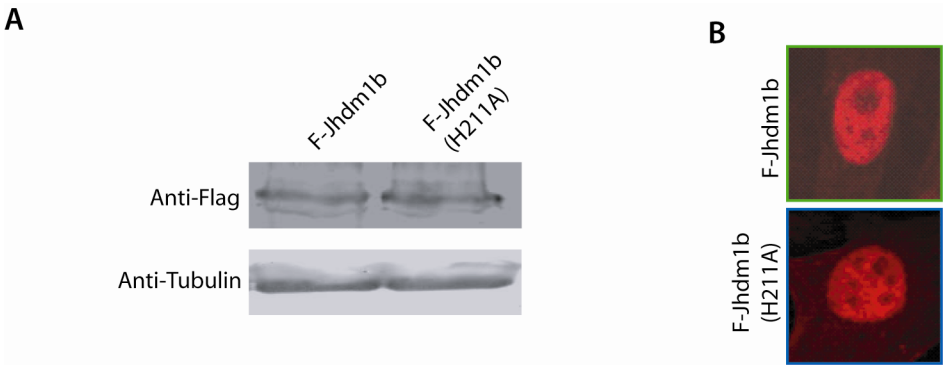

**Table 3-1.** All primers sequences used in the experiments described in Chapter 3.

| Figure# | Name | Use | Sequence 5'-3' |
|---|---|---|---|
| 3-1D | hmJHDM1B-5 | RT-PCR | GACTTGTCGGACGTGGAGGA |
| 3-1D | hmJHDM1B-3 | RT-PCR | CACATGTTGTCCACCCAGTC |
| 3-1D, 3-2A, 3-3A | GAPDH5 | RT-PCR | CATGGCCTTCCGTGTTCCTA |
| 3-1D, 3-2A, 3-3A | GAPDH3 | RT-PCR | GCCTGCTTCACCACCTTCTT |
| 3-2A, 3-3A | JHDM1b-5 | RT-PCR | ACGTGATCCGGCCACCGCCCATC |
| 3-2A, 3-3A | JHDM1b-3 | RT-PCR | TTGGAGATGTTGGTCCAACTGAG |
| 3-2A, 3-3A | Ring1b-5 | RT-PCR | TAGTGGAGCAGAAGATAATGG |
| 3-2A, 3-3A | Ring1b-3 | RT-PCR | TCAACAGTGGCATTGCCTGAAG |
| 3-3A, 3-4B | P15INK4b-5 | RT-PCR | ATGTTGGGCGGCAGCAGTGACG |
| 3-3A, 3-4B | P15INK4b-3 | RT-PCR | ATCTCCAGTGGCAGCGTGCAG |
| 3-3A | P16INK4a-5 | RT-PCR | TACCCCGATTCAGGTGATGATG |
| 3-3A | P16INK4b-3 | RT-PCR | TAGCTCTGCTCTTGGGATTGG |
| 3-3A | P19ARF-5 | RT-PCR | TGATGTTTGGAAGTCCAGCAG |
| 3-3A | P19ARF-3 | RT-PCR | AATGTCCATGAGGTTCTGAGC |
| 3-3A | P18-5 | RT-PCR | AACGTCAACGCTCAAAATGG |
| 3-3A | P18-3 | RT-PCR | AGGCTGTGTGCTTCATAAGG |
| 3-3A | P21-5 | RT-PCR | ATGTCCAATCCTGGTGATGTC |
| 3-3A | P21-3 | RT-PCR | TCTCTTGCAGAQAGACCAATC |
| 3-3A | P27-5 | RT-PCR | TGAGAGTGTCTAACGGGAGCC |
| 3-3A | P27-3 | RT-PCR | TCTGACGAGTCAGGCATTTGGTC |
| 3-4D, E | mink4b1-5 | ChIP | CCGCCTAGAGATCGAACTAGCC |
| 3-4D, E | mink4b1-3 | ChIP | CGCTTTTGCAATTGACTGAC |
| 3-4D, E | mink4b2-5 | ChIP | CACCGAAGCTACTGGGTCTC |

| | | | |
|---|---|---|---|
| 3-4D, E | mink4b2-3 | ChIP | CTGTGGCAGAAATGGTCCTT |
| 3-4D, E | mink4b3-5 | ChIP | ATGTTCTAAGAGGCTTTGTTTCCA |
| 3-4D, E | mink4b3-3 | ChIP | CATTTGTGCATAGGAGATCAGG |
| 3-4E | mgapdh-5 | ChIP | CCACCCAGAAGACTGTGGAT |
| 3-4E | mgapdh-3 | ChIP | GGATGCAGGGATGATGTTCT |

## REFERENCES


Bannister, A. J., R. Schneider, et al. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem* **280**(18): 17732-6.

Bracken, A. P., D. Kleine-Kohlbrecher, et al. (2007). The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes Dev* **21**(5): 525-30.

Cales, C., M. Roman-Trufero, et al. (2008). Inactivation of the polycomb group protein Ring1B unveils an antiproliferative role in hematopoietic cell expansion and cooperation with tumorigenesis associated with Ink4a deletion. *Mol Cell Biol* **28**(3): 1018-28.

Cao, R., Y. Tsukada, et al. (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol Cell* **20**(6): 845-54.

Cao, R., H. Wang, et al. (2008). Role of hPHF1 in H3K27 methylation and Hox gene silencing. *Mol Cell Biol* **28**(5): 1862-72.

Frescas, D., D. Guardavaccaro, et al. (2007). JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes. *Nature* **450**(7167): 309-13.

Gil, J. and G. Peters (2006). Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. *Nat Rev Mol Cell Biol* **7**(9): 667-77.

Goodrich, D. W., N. P. Wang, et al. (1991). The retinoblastoma gene product regulates progression through the G1 phase of the cell cycle. *Cell* **67**(2): 293-302.

Jacobs, J. J., K. Kieboom, et al. (1999). The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. *Nature* **397**(6715): 164-8.

Kamb, A., N. A. Gruis, et al. (1994). A cell cycle regulator potentially involved in genesis of many tumor types. *Science* **264**(5157): 436-40.

Kannengiesser, C., S. Dalle, et al. (2007). New founder germline mutations of CDKN2A in melanoma-prone families and multiple primary melanoma development in a patient receiving levodopa treatment. *Genes Chromosomes Cancer* **46**(8): 751-60.

Klose, R. J., Q. Yan, et al. (2007). The retinoblastoma binding protein RBP2 is an H3K4 demethylase. *Cell* **128**(5): 889-900.

Kotake, Y., R. Cao, et al. (2007). pRB family proteins are required for H3K27 trimethylation and Polycomb repression complexes binding to and silencing p16INK4alpha tumor suppressor gene. *Genes Dev* **21**(1): 49-54.

Koyama-Nasu, R., G. David, et al. (2007). The F-box protein Fbl10 is a novel transcriptional repressor of c-Jun. *Nat Cell Biol* **9**(9): 1074-80.

Krimpenfort, P., A. Ijpenberg, et al. (2007). p15Ink4b is a critical tumour suppressor in the absence of p16Ink4a. *Nature* **448**(7156): 943-6.

Martin, C. and Y. Zhang (2005). The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* **6**(11): 838-49.

Morris, S. A., Y. Shibata, et al. (2005). Histone H3 K36 methylation is associated with transcription elongation in Schizosaccharomyces pombe. *Eukaryot Cell* **4**(8): 1446-54.

Ogawa, S., N. Hirano, et al. (1994). Homozygous loss of the cyclin-dependent kinase 4-inhibitor (p16) gene in human leukemias. *Blood* **84**(8): 2431-5.

Okuda, T., S. A. Shurtleff, et al. (1995). Frequent deletion of p16INK4a/MTS1 and p15INK4b/MTS2 in pediatric acute lymphoblastic leukemia. *Blood* **85**(9): 2321-30.

Ortega, S., I. Prieto, et al. (2003). Cyclin-dependent kinase 2 is essential for meiosis but not for mitotic cell division in mice. *Nat Genet* **35**(1): 25-31.

Pfau, R., A. Tzatsos, et al. (2008). Members of a family of JmjC domain-containing oncoproteins immortalize embryonic fibroblasts via a JmjC domain-dependent process. *Proc Natl Acad Sci U S A* **105**(6): 1907-12.

Rao, B., Y. Shibata, et al. (2005). Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Mol Cell Biol* **25**(21): 9447-59.

Rosu-Myles, M., B. J. Taylor, et al. (2007). Loss of the tumor suppressor p15Ink4b enhances myeloid progenitor formation from common myeloid progenitors. *Exp Hematol* **35**(3): 394-406.

Sherr, C. J. and J. M. Roberts (1999). CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev* **13**(12): 1501-12.

Suzuki, T., K. Minehata, et al. (2006). Tumor suppressor gene identification using retroviral insertional mutagenesis in Blm-deficient mice. *Embo J* **25**(14): 3422-31.

Tsukada, Y., J. Fang, et al. (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**(7078): 811-6.

Voncken, J. W., B. A. Roelen, et al. (2003). Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc Natl Acad Sci U S A* **100**(5): 2468-73.

Wang, H., L. Wang, et al. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**(7010): 873-8.

Xiao, B., C. Jing, et al. (2003). Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* **421**(6923): 652-6.

Zeng, P. Y., C. R. Vakoc, et al. (2006). In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. *Biotechniques* **41**(6): 694, 696, 698.

Zhang, X., H. Tamaru, et al. (2002). Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell* **111**(1): 117-27.

# Chapter 4

## Genome wide uH2A localization analysis highlights Bmi1-dependent deposition of the mark at repressed genes

**ABSTRACT**

Polycomb group (PcG) proteins control organism development by regulating the expression of developmental genes. Transcriptional regulation by PcG proteins is achieved at least partly through the PRC2-mediated methylation on lysine 27 of histone H3 (H3K27) and PRC1-mediated ubiquitylation on lysine 119 of histone H2A (uH2A). As an integral component of PRC1, Bmi1 has been demonstrated to be critical for H2A ubiquitylation. Although recent studies have revealed the genome wide binding patterns of some of the PRC1 and PRC2 components, as well as the H3K27me3 mark, there have been no reports describing genome wide localization of uH2A. Using the recently developed ChIP-Seq technology, here we report genome wide localization of the Bmi1-dependent uH2A mark in MEF cells. Gene promoter averaging analysis indicates a peak of uH2A just inside the transcription start site (TSS) of well annotated genes. This peak is enriched at promoters containing the H3K27me3 mark and represents the least expressed genes in WT MEF cells. In addition, peak finding reveals regions of local uH2A enrichment throughout the mouse genome, including almost 700 gene promoters. Genes with promoter peaks of uH2A exhibit lower level expression when compared to genes that do not contain promoter peaks of uH2A. Moreover, we demonstrate that genes with uH2A peaks have increased expression upon Bmi1 knockout. Importantly, local enrichment of uH2A is not limited to regions containing the H3K27me3 mark. We describe the enrichment of H2A ubiquitylation at high density CpG promoters and provide evidence to suggest that DNA methylation may be linked to uH2A at these regions. Thus, our work not only reveals Bmi1-dependent H2A ubiquitylation but also suggests that uH2A targeting in differentiated cells may employ a different mechanism from that in ES cells.

## INTRODUCTION

In higher eukaryotes, DNA is organized in the form of chromatin. The basic repeating unit of chromatin is called the nucleosome, which consists of 146 bp of DNA wrapped around a core histone octamer. One unique feature of core histones is their proclivity for covalent modification including acetylation, methylation, ubiquitylation, and phosphorylation (Jenuwein and Allis 2001). In addition, DNA can be modified directly through methylation. These covalent modifications can affect gene transcription directly or indirectly through the recruitment of additional modulatory factors (Martin and Zhang 2005). Therefore, different combinations of modifications on chromatin may ultimately determine distinct cellular states through regulating the transcriptional programs that cells adopt. Thus, identification and characterization of the proteins that are responsible for the placement and maintenance of these epigenetic marks is of great importance in understanding cellular proliferation and differentiation.

The addition of a single ubiquitin molecule to histone H2A at lysine 119 was first discovered over thirty years ago (Olson, Goldknopf et al. 1976). Classic experiments demonstrated that uH2A accounts for about 10% of total H2A (Levinger and Varshavsky 1980). Despite the knowledge of its existence, the identity of the responsible enzymes and the function of this modification have only recently begun to be elucidated. The first H2A ubiquitin E3 ligase was identified as the core components of the Polycomb repressive complex 1 (PRC1) composed of RING1/2, BMI1, and HPH2 (Wang, Wang et al. 2004). Biochemical and functional analysis of the PRC1 complex has revealed RING2/Ring1b as the catalytic subunit, which can be greatly stimulated by Bmi1 and Ring1a, as loss function on any of these two proteins resulted in drastic genome-wide reduction of uH2A (Cao,

Tsukada et al. 2005; Buchwald, van der Stoop et al. 2006). Genome-wide location studies revealed that PRC1 occupies the promoters of a subset of Polycomb repressive complex 2 (PRC2) targets and both PRC1 and PRC2 are enriched at genes involved in developmental processes (Boyer, Plath et al. 2006; Bracken, Dietrich et al. 2006; Lee, Jenner et al. 2006). Recent studies have uncovered that Bmi1 homologs, such as Mel18 and NSPc1, can target the PRC1 complex in various cell types (Elderkin, Maertens et al. 2007; Wu, Gong et al. 2008). In addition, a new E3 ligase for H2A, 2A-HUB, has also been reported (Zhou, Zhu et al. 2008) highlighting the fact that there must be Bmi1-dependent and Bmi1-independent pools of uH2A in the genome.

Unlike PRC1, PRC2 possesses H3K27-specific histone methyltransferase activity (Cao and Zhang 2004). The discovery that a component of PRC1, Pc, can specifically recognize and bind to H3K27me3 (Cao, Wang et al. 2002; Fischle, Wang et al. 2003; Min, Zhang et al. 2003) has prompted researchers to embrace a sequential recruitment model whereby PRC2-mediated H3K27 methylation contributes to PRC1 recruitment and subsequent ubiquitylation of histone H2A. This model is supported by three pieces of evidence. First, studies on Hox and *Ink4a/Arf* loci indicate that PRC1 knockdown reduced local uH2A levels which correlate with upregulation of gene expression (Jacobs, Kieboom et al. 1999; Cao, Tsukada et al. 2005; Kotake, Cao et al. 2007). Second, knockdown of the H3K27me3 demethylase, Utx, results in enrichment of both PRC1 and uH2A at PRC2 target genes (Lee, Villa et al. 2007). Third, the majority of genome-wide Ring1b enriched regions in mouse embryonic stem cells (mES) co-localize with peaks of H3K27me3 (Ku, Koche et al. 2008).

In addition to the relationship between H3K27 methylation and H2A ubiquitylation, several studies also suggest a potential link between H3K27 methylation and DNA methylation. For example, H3K27 methylation has been demonstrated to play an important role in imprinted gene silencing (Lewis, Mitsuya et al. 2004; Umlauf, Goto et al. 2004). Components of PRC2, such as Ezh2, have been reported to interact with Dnmt1/3a/3b and are required for efficient DNA methylation at several target genes (Vire, Brenner et al. 2006). On the other hand, Dnmt1 may contribute to the recruitment of PRC1 as knockdown of Dnmt1 abrogates localization of PRC1 components to Polycomb bodies in cultured cells (Hernandez-Munoz, Taghavi et al. 2005). Consistent with this notion, recent studies have demonstrated that components of PRC1 can interact with a methyl-DNA binding protein, Mbd1 (Sakamoto, Watanabe et al. 2007), and the Dnmt1-associated protein, Dmap1 (Negishi, Saraya et al. 2007). Despite these reports, a general correlation between H3K27 methylation and DNA methylation may not exist as genome wide epigenetic profiling revealed only a small subset of H3K27me3 positive promoters were found to be hypermethylated (Fouse, Shen et al. 2008; Kondo, Shen et al. 2008). Whether there exists a genome-wide link between PRC1 mediated H2A ubiquitylation and DNA methylation remains to be determined.

The advent of chromatin immunoprecipitation coupled to genomic tiling arrays (ChIP-chip) has provided scores of reports highlighting genome wide maps of histone modifications (Barrera and Ren 2006), histone modifying enzyme binding profiles (Bracken, Dietrich et al. 2006), and transcription modulators (Kim, Abdullaev et al. 2007; Xu, Bieda et al. 2007). Recent advances in ChIP-coupled deep sequencing have greatly expedited the tedious task of dissecting the interplay between epigenetic modifications and complex

transcriptional output (Schones and Zhao 2008). Although genome-wide analysis of most epigenetic marks (Barski, Cuddapah et al. 2007; Mikkelsen, Ku et al. 2007; Wang, Zang et al. 2008), as well as transcription factors (Chen, Xu et al. 2008), has been reported for various cell lines, uH2A distribution remains a mystery. In addition, the recent discovery that the majority of uH2A in the fly genome is placed by a complex containing the Bmi1 homolog but lacking Pc (Lagarou, Mohd-Sarip et al. 2008) calls the generality of the sequential recruitment model into question.

To understand how uH2A fits into the complex epigenetic architecture associated with mammalian chromatin, we describe the genome wide profile of Bmi1-dependent uH2A by comparing the enrichment of this mark in Bmi1 wild-type and null MEF cells. This analysis provides evidence that while Bmi1 dependent uH2A is enriched at genes containing the H3K27me3 mark, it is not limited to these regions. In addition, analysis of genome-wide DNA methylation patterns reveals a link between uH2A and DNA methylation in high-density CpG promoters. Transcription profiling of wild-type MEF cells indicates that Bmi1-dependent uH2A is enriched at genes with low levels of *de novo* expression. Finally, genes containing the highest levels of Bmi1 dependent uH2A at their promoters are expressed higher upon Bmi1 loss of function than genes harboring low levels of uH2A. Thus, our study uncovers some previously unrecognized features of uH2A.

## RESULTS

*Generating genome wide uH2A modification maps*

In an effort to understand the function of the Bmi1-dependent uH2A epigenetic mark, we performed chromatin immunoprecipitation (ChIP) experiments using the well characterized uH2A monoclonal antibody E6C5 (Vassilev, Rasmussen et al. 1995; de Napoles, Mermoud et al. 2004; Wang, Wang et al. 2004; Cao, Tsukada et al. 2005; Stock, Giadrossi et al. 2007) in wild-type MEFs. The precipitated DNA was subjected to deep sequencing using the Solexa sequencing technology. As a control for non-specific background, and to focus the research on the Bmi1-dependent proportion of genomic uH2A, parallel ChIPs were also performed in Bmi1 null MEF cells. A previous report has shown that these cells undergo drastic reductions in global H2A ubiquitylation (Cao, Tsukada et al. 2005). After sequencing 25 bp DNA fragments, the data retrieved from both cell types were mapped to the *mm*8 build of the *Mus musculus* genome. DNA tags which did not uniquely map to the genome were discarded and the resulting tag libraries consisted of over 6 million and 8 million unique reads for the wild-type and Bmi1 null MEF cells, respectively. We next performed normalization for the total number of uniquely mapped reads, and generated a final Bmi1-dependent uH2A summary file by aligning reads from both libraries and subtracting uH2A tags derived from Bmi1 null MEFs from those derived from wild-type MEFs. Density maps were created by ignoring negative tag density regions (data not shown).

*Enrichment of uH2A at the majority of gene promoters is dependent on Bmi1 function*

Given the enrichment of many epigenetic marks at gene promoters, we began our analysis of Bmi1-dependent uH2A localization by examining the area directly surrounding

the transcription start site of well annotated genes. To this end, the average per base pair normalized density of Bmi1-dependent uH2A was determined for a 10 kb region surrounding well annotated TSSs at a 200 bp resolution. This genome wide averaging analysis revealed enrichment of the uH2A mark that peaked just inside the TSS (Fig. 4-1A). Given the link between uH2A and PRC2 (Hernandez-Munoz, Taghavi et al. 2005), we next set out to determine if a quantitative relationship existed between Bmi1-dependent promoter uH2A enrichment and promoter H3K4me3 and/or H3K27me3 placement. To this end, well-annotated genes were grouped into four classes based on the presence of H3K4me3 and H3K27me3 within their promoters, as determined by a previous study (Mikkelsen, Ku et al. 2007). The same averaging analysis was applied to these separate groups of genes which included promoters containing H3K4me3 only, H3K27me3 only, both H3K4me3 and H3K27me3 (bivalent), or lacking both modifications (no K4/no K27). This analysis revealed Bmi1-dependent enrichment of average uH2A tag density for both the bivalent and H3K27me3 modified gene classes (Fig. 4-1B, compare grey line to green and blue lines). Enrichment manifested in an increase in averaged peak height, as well as an overall broadening of the uH2A peak further into the body of K27me3 marked genes. This result is consistent with a role for PRC2 in the recruitment of PRC1 and subsequent ubiquitylation of histone H2A (Wang, Brown et al. 2004; Cao, Tsukada et al. 2005). Interestingly, Bmi1-dependent uH2A was still present at H3K4me3 genes at a level comparable to the all gene average (Fig. 4-1B, red line). Furthermore, genes lacking both H3K4me3 and H3K27me3 exhibited a depletion of uH2A tag density (Fig. 4-1B, purple line). Together these gene region averaging results reveal an overall enrichment of Bmi1-dependent uH2A at gene promoters which is biased towards genes marked by H3K27me3.

*Bmi1-dependent uH2A are enriched at specific genomic regions*

To complement the genome wide averaging studies described above, the uH2A tag library of wild-type MEFs was subjected to peak enrichment analysis using the TIROE program (Ho, Jothi et al. 2009) with the uH2A tag library derived from Bmi1 null MEFs set as background. This data processing allowed us to identify the most enriched Bmi1-dependent peaks of uH2A throughout the genome.

The TIROE program identified 16,406 peaks of Bmi1-dependent uH2A throughout the genome with enrichment *P* values of at least 1e-5. We began our analysis of peak distribution by roughly dividing the genome into genic and non-genic regions as determined by the presence of transcribed regions annotated in the REFSEQ database. Interestingly, while genic regions correspond to only about 47% of the genome, they harbored 52% of the identified peaks (Table 4-1). Further division of the genic regions of the mouse genome into promoter and transcribed regions revealed that peak number enrichment was present in both of these sub-groups with 671 peaks (4.1%) falling within gene promoters and the remaining 7415 peaks (47.3%) localizing elsewhere along the transcribed body of the gene (Table 4-1). Distribution analysis of defined peaks along transcribed regions revealed an increased tendency of uH2A peak localization towards the transcription termination site of genes (Fig. S4-1A). However, in agreement with tag library averaging data presented in Figure 4-1A, peaks that fell within the promoter of genes (defined as -1 kb to +1 kb around TSSs) exhibited an average tag density higher than those lying within gene bodies (Fig. S4-1B). Taken together, these data indicate that peaks of Bmi1-dependent uH2A are enriched within the transcribed regions of well-annotated genes with smaller peaks clustered towards the transcription termination site and larger peaks specifically occupying gene promoters.

Given the enrichment of tag density at gene promoters as visualized by both genome-wide averaging (Fig. 4-1A) and peak localization (Table 4-1 & Fig. S4-1B), we next selected several representative TIROE peaks that localized within gene promoters to be verified by ChIP-qPCR. ChIP assays were carried out in both wild-type and Bmi1 null MEFs using antibodies against uH2A and Bmi1. The Bmi1 null subtracted tag density profiles surrounding the TSS of *Cebpa*, *B4galnt1*, *Gfod2*, *Zfp12*, *Fgf6*, *Dcxr*, *Iars*, *Arpc3* and *Chmp2a* are shown in Figure 2A. These peaks have TIROE enrichment *P* values of 2.4e-15, 1.2e-14, 1.3e-12, 7.0e-11, 9.2e-09, 1.1e-08, 1.1e-07, 2.2e-06, and 3.0e-06, respectively. ChIP assays not only confirmed the enrichment of uH2A at these loci in wild-type cells as compared to Bmi1 null cells, but also the presence of Bmi1 at these same regions in wild-type cells (Fig. 4-2B). Similar enrichment of uH2A and Bmi1 localization was not detected within the promoters of *Mta2*, *Hbb2*, and *Ppara*; genes which do not contain peaks of the uH2A epigenetic mark (Fig. 4-2B).

*Peaks of Bmi1-dependent uH2A overlap with other epigenetic marks*

Genetic studies have revealed a link between PRC1-catalyzed H2A ubiquitylation and PRC2-catalyzed H3K27 methylation. Biochemically, H3K27me3 has been shown to serve as a binding site for the recruitment of the PRC1 complex (Fischle, Wang et al. 2003; Min, Zhang et al. 2003). These studies suggest that there should be a link between uH2A and H3K27me3. However, recent evidence from *Drosophila* has implicated another Bmi1 containing complex, dRAF, in H2A ubiquitylation. Interestingly, this complex lacks Pc and must utilize an alternative means of targeting (Lagarou, Mohd-Sarip et al. 2008). To examine whether our dataset could shed light on the relationship between the two modifications, we investigated the overlap between peaks of Bmi1-dependent uH2A,

H3K27me3, and H3K4me3. We reasoned that if an alternate method of E3 ligase recruitment was present in MEF cells then uH2A enrichment would be present in regions not marked by H3K27me3. Of the 4132 peaks of H3K27me3 identified in a previous study (Mikkelsen, Ku et al. 2007), about 15% (651) overlap with peaks of uH2A. Furthermore, about 11% of the 2,604 genomic regions defined as bivalent overlap with our uH2A dataset. Interestingly, about 5.6% of the 14,178 peaks of H3K4me3 exhibit co-localization with peaks of Bmi1-dependent uH2A. Together, these findings reveal that enriched regions of uH2A show a marked localization bias to genomic regions which also contain the H3K27me3 mark. However, they also indicate that the vast majority of Bmi1-dependent uH2A falls outside of regions containing this mark and suggests that an alternate method of ubiquitin E3 ligase recruitment may exist in MEF cells.

Given that about half of all uH2A enriched regions in the genome fall outside of genic regions and only 4% are localized to gene promoters (Table 4-1), we next explored possible differences between promoter and non-promoter peaks of uH2A with regards to H3K4me3 and H3K27me3 overlap. To this end, peaks were binned by their localization inside or outside of promoters, and the number of peaks from each of these groups which overlapped with additional epigenetic modifications was calculated. This analysis revealed that a higher proportion of promoter bound Bmi1-dependent uH2A peaks co-localized with either H3K4me3 or H3K27me3 when compared to peaks located outside of promoter regions (Fig. 4-3A). This result is not surprising given the relative enrichment of H3K4/H3K27 within gene promoter regions. However, this finding reinforces the fact that Bmi1-dependent uH2A is distinct from the H3K27me3 mark with respect to the extent of its enrichment outside of genic regions.

We next investigated the local tag density of uH2A peaks both within and outside of promoters. Results presented in Figure 4-3B show that on average, peaks of uH2A localized within gene promoters are composed of significantly more tags than peaks lying outside of gene promoters (Wilcoxon *P* value = 2.2e-3). In addition, further sub-division of promoter and non-promoter peaks by H3K4me3 and H3K27me3 overlap reveals two additional pieces of information. First, promoter specific enrichment of uH2A peak tag density occurs at genes that also contain the H3K27me3 mark (Fig 4-3C, compare promoter peak data sets). This result is consistent with whole genome averaging analysis presented in Figure 4-1B. Second, there is a depletion of average peak tag density at non-promoter regions that do not contain either H3K4me3 or H3K27me3 (Fig 4-3C, compare non-promoter peak data sets). This finding, in conjunction with the fact that the vast majority of non-promoter uH2A falls outside of H3K4me3 and H3K27me3 regions (Fig. 4-3A), suggests that the contribution of these independent peaks is driving overall depletion of tag density at non-promoter regions.

*Promoter bound uH2A is genetically linked to DNA methylation*

Previous studies indicate that PRC2 facilitates recruitment of DNA methyltransferases to at least a sub-set of Polycomb target genes (Vire, Brenner et al. 2006). This finding, together with the relationship between DNA methylation and transcriptional repression prompted us to investigate a possible link between genome-wide promoter uH2A and DNA methylation. To this end, we first analyzed uH2A localization in relation to CpG dinucleotide content surrounding the TSS of known genes. Following a previous characterization (Mikkelsen, Ku et al. 2007), we divided the genes into three groups defined as high level (*n* = 10,310), intermediate level (*n* = 2889), and low level (*n* = 2668) based on the density of CpGs within their promoter regions (HCP, ICP, and LCP, respectively). We

repeated whole genome promoter averaging analysis and found that the promoter tag density of uH2A is strongly enriched at genes defined as HCP when compared to both ICP and LCP genes (Fig. 4-4A). Recent reports have used bisulphite treatment coupled with deep sequencing to characterize the extent of DNA methylation at HCP group genes in ES, NPC, and MEFs (Meissner, Mikkelsen et al. 2008). Comparison of the available DNA methylation data with our promoter uH2A peaks revealed a correlation between DNA methylation levels and average uH2A tag density in the HCP group (Fig. 4-4B). Specifically, HCP genes with promoter peaks of Bmi1-dependent uH2A had higher levels of average DNA methylation when compared to HCP genes without uH2A peaks (Wilcoxon *P* value = 0.0426).

To further characterize this correlation we asked whether the two modifications are genetically connected. Toward this end, we asked whether loss of DNA methylation would cause alteration in H2A ubiquitylation. A comparison of the uH2A levels in Dnmt1 null MEFs (Jackson-Grusby, Beard et al. 2001) revealed that loss of DNA methylation resulted in a significant decrease in uH2A levels (Fig. 4-4C). To determine whether alteration in uH2A level affects DNA methylation, we compared the DNA methylation levels in wild-type and Bmi1 null MEFs by cytosine extension analysis. This technique takes advantage of DNA methylation sensitive/insensitive restriction enzymes and allows for the relative quantification of DNA methylation through end-labeling of genomic DNA digestion products (Fujiwara and Ito 2002). Results shown in Figures 4-4D and 4-4E revealed a small, but statistically insignificant difference in the DNA methylation levels in the wild-type and Bmi1 null MEFs as determined by Student's t-test. As expected, parallel analysis revealed a drastic decrease of DNA methylation in the Dnmt1 null MEFs as indicated by the increased sensitivity of genomic DNA to the methylation sensitive restriction enzyme, *HpaII* (Fig. 4-

4D, E). Taken together, these data not only revealed a correlation between promoter bound uH2A and HCP promoter DNA methylation, but also provide evidence that DNA methylation may be upstream of H2A ubiquitylation.

*Bmi1-dependent promoter uH2A marks repressed genes through its co-localization with the H3K27me3 mark*

Recent studies suggest a role for uH2A in the repression of developmental genes (Cao, Tsukada et al. 2005; Bracken, Dietrich et al. 2006; Kim, Paylor et al. 2006). In addition, studies utilizing *in vitro* assembled chromatin templates have implicated uH2A in the repression of transcription initiation (Nakagawa, Kajitani et al. 2008). To determine whether Bmi1-dependent uH2A is a general indicator of transcriptional activity, we asked whether uH2A levels and gene expression levels have a general correlation. To this end, we profiled gene expression in wild-type MEFs using the Affymetrix Mouse Genome 430 2.0 microarray and analyzed the relationship between presence of uH2A peaks and gene activity for 13,354 genes. Based on their expression level, genes were grouped into 10 equal sized bins and were correlated with promoter uH2A enrichment levels using genome-wide tag density averaging. This analysis revealed that while the peak height of uH2A enrichment did not change much over these expression groups, the lowest expressed gene groups exhibited a broadening of the average uH2A tag density peak into the body of genes (Fig. 4-5A). To better visualize this trend, the region spanning from +0.6 kb to +2.0 kb of genes was reanalyzed with genes divided into 4 groups. The resulting plot (Fig. 4-5B) confirms Bmi1-dependent tag density broadening for genes present within the lowest expressed gene groups, a medium range broadening for genes in the third group, and no tag density broadening for the highest expressed genes present within the fourth group.

We next turned our attention to distinct promoter peaks of uH2A in an effort to understand gene specific outcomes as related to epigenetic mark deposition. In agreement with the data described above, expression level averaging of genes containing promoter peaks of uH2A revealed them to be significantly lower expressed then those without uH2A peaks (Wilcoxon *P* value = 9.8e-3) (Fig. 4-5C). Taken together, these data indicate that an overall increase in Bmi1-dependent uH2A abundance can be found at silenced or low expressed genes; likewise, genes marked by promoter peaks of uH2A tend to be expressed at a lower level then genes lacking peaks of Bmi1-dependent uH2A.

The correlation between increased levels of uH2A and gene repression could be due to an intrinsic repressive effect of uH2A on transcription or due to an association of uH2A with other silencing epigenetic marks. To differentiate between these possibilities, we compared genes with promoter uH2A enrichment peaks to those without peaks in terms of their association with promoter H3K4me3 or H3K27me3 marks. This analysis revealed that a higher proportion of genes containing uH2A are also marked by H3K27me3 either alone or in the context of bivalent domains when compared with genes not marked by uH2A (Fig. 4-5D). In addition, the uH2A positive gene set is enriched for promoters lacking both H3K27me3 and H3K4me3 as well as depleted for the H3K4me3 mark alone (Fig. 4-5D). This indicates that grouping genes based on uH2A results in a population enriched for epigenetic marks tightly linked to gene repression and depletion of epigenetic marks tightly linked to gene expression. Given that this proportional shift could indicate that uH2A is passively correlated to gene repression, we next asked the question whether uH2A could function in conjunction with these additional epigenetic marks to fine-tune mRNA expression levels. To this end, we repeated the analysis described in Figure 4-5C for uH2A

gene groups further sub-divided by additional promoter modifications and compared the average expression of these gene groups across classes (Fig. 4-5E). Results confirm higher average expression of genes marked by H3K4me3 and drastically lower expression levels of genes containing promoter H3K27me3, H3K27me3 and H3K4me3 (bivalent), and neither mark (Fig. 4-5E). Co-occurrence of uH2A with H3K4me3, H3K27me3, or bivalent domains did not result in any changes of average expression level when compared to genes harboring only the K4/K27 combinations. The sole significant change in WT expression level was a small increase for genes containing only the uH2A mark (Wilcoxon *P* value = 0.026) (Fig. 4-5E). These results indicate that the correlation that exists between promoters bound by Bmi1-dependent uH2A and low level *de novo* gene expression is due to a proportional shift in other epigenetic marks that have a more profound effect on transcriptional state.

Even though Bmi1-dependent promoter uH2A is enriched at silenced genes, results of *de novo* expression analysis in wild-type cells were not able to ascertain an active function for uH2A in gene silencing. It is possible that a global role for uH2A in gene silencing was masked in this analysis by the presence of other epigenetic modifications which are more potent in enacting transcriptional control. To directly address a potential global role for Bmi1-dependent uH2A in gene silencing, we turned our attention to expression changes upon Bmi1 knock-out. To this end, we performed a microarray study on Bmi1 null MEFs and compared gene expression with that in the wild-type MEFs. Of the 671 genes marked by promoter uH2A, we were able to generate reliable fold-change transcription data for 472 genes. We found that on average, this group was upregulated in Bmi1 null MEF cells by 1.5 fold. We next investigated whether the level of uH2A loss at genes in Bmi1 null cells was correlated to increased expression upon Bmi1 knock-out. For this analysis, genes containing

Bmi1-dependent peaks of uH2A were sorted into three groups by increasing tag density ($n$ = 157, 157, and 158, respectively), and both the average fold change as well as the data distribution of each group was determined. Interestingly, genes marked by the lowest levels of uH2A exhibited the smallest average increase in transcription upon Bmi1 knock-out when compared to both intermediate and high tag density groups (1.14, 1.98, and 1.37 fold, respectively). In addition, data distribution analysis and Wilcoxon signed-rank testing reveals that this finding is statistically significant (Fig. 5F). Taken together, these data are consistent with a global role for Bmi1-dependent uH2A in gene silencing and extend gene specific analysis at important developmental regulators to genome-wide correlation.

## DISCUSSION

Deep sequencing techniques have recently been used to map epigenetic marks in high definition throughout mammalian genomes (Barski, Cuddapah et al. 2007; Mikkelsen, Ku et al. 2007; Meissner, Mikkelsen et al. 2008). Even though mono ubiquitylation of H2A was one of the first histone modifications identified (Olson, Goldknopf et al. 1976), it remains among the least understood. Using a ChIP-seq approach, here we analyzed the genome-wide distribution of Bmi1-dependent uH2A. This investigation revealed several interesting features of this epigenetic modification which serve as the basis for further studies.

*Bmi1-dependent uH2A distribution throughout mammalian chromatin*

By combining genome averaging and peak localization analyses, this study reveals the first picture of the genome-wide localization of the uH2A mark and identifies Bmi1-dependent enrichment within both genic and non-genic regions of the mouse genome. On average, uH2A tag density is enriched at gene promoter regions (Fig. 4-1A) with further enrichment at genes marked by H3K27me3 (Fig. 4-1B). Peak analysis reveals that Bmi1-dependent uH2A enriched regions coincide with gene promoters at a much higher rate than can be expected by chance (Table 4-1), and these peaks encompass significantly higher tag values when compared with non-promoter peaks (Fig. 4-3B). In addition, promoter peak tag values are enriched at genes also marked by H3K27me3 (Fig. 4-3C). Interestingly, the gross distribution of genic uH2A peaks is skewed towards the 3' end of genes, indicating that more regions of uH2A enrichment are found within these genic regions (Fig. S4-1A). However, these peaks represent lower enrichment regions when compared to promoter peaks (Fig. S4-1B). Together, these data support the notion that H3K27me3 can contribute to the recruitment of PRC2 and ubiquitylation of H2A in promoters (Fischle, Wang et al. 2003;

Min, Zhang et al. 2003). However, even though the highest regions of Bmi- dependent uH2A enrichment are at H3K27me3 genes, clear peaks also exist at genomic regions (both genic and non-genic) that do not contain this mark. This uH2A distribution pattern is very different from what would be expected given a recent study of Ring1b and Bmi1 binding profiles in mES cells (Ku, Koche et al. 2008) but are more consistent with a previous study which described a proportion of gene promoters positive for Ring1b/Bmi1 but negative for PRC2 binding (Boyer, Plath et al. 2006). Although this discrepancy may reflect the differences of uH2A distribution in ES cells and MEF cells, it is possible that similar to the observations in *Drosophila* (Lagarou, Mohd-Sarip et al. 2008), a Pc independent mechanism of Ring1b/Bmi1 recruitment may exist in MEF cells. Along these lines, about half of all the most enriched Bmi1-dependent uH2A regions lie outside of both promoter and transcribed regions of the genome (Table 4-1). Interestingly, these regions do not overlap significantly with peaks of H3K27me3.

*Relationship between uH2A and DNA methylation*

Since the initial finding linking Polycomb silencing to DNA methylation (Vire, Brenner et al. 2006), genome wide studies have called the generality of this association into question as very little overlap exists between genes methylated at H3K27 and genes that contain high levels of CpG island methylation associated with their promoters (Farthing, Ficz et al. 2008; Fouse, Shen et al. 2008; Kondo, Shen et al. 2008). Recently, several reports have demonstrated a role for PRC1 in recognizing methylated DNA at specific loci or heterochromatic regions (Hernandez-Munoz, Taghavi et al. 2005) either through Bmi1 interaction with Dmap1 (Negishi, Saraya et al. 2007) or Ring1b interaction with Mbd1 (Sakamoto, Watanabe et al. 2007). However, whether these specific instances of

convergence of silencing pathways are linked to global gene regulation has not been determined. Here we provide evidence to support a functional link between DNA methylation and histone ubiquitylation through Bmi1-dependent mechanisms in a group of high CpG-containing genes. Not only does the level of DNA methylation increase at HCP promoters marked by uH2A (Fig. 4-4B), a result that may be explained by an increase in H3K27 methylation on these genes, but knock-out of Dnmt1, which results in the ablation of CpG methylation genome wide, causes a global decrease in the uH2A levels (Fig. 4-4C). In contrast, Bmi1 knock-out results in only a small, statistically insignificant, increase in global DNA methylation (Fig. 4-4D). Taken together, these data suggest that DNA methylation may be upstream of Bmi1-dependent H2A ubiquitylation. Future work will reveal how DNA methylation contributes to H2A ubiquitylation.

*The relationship between uH2A and gene expression*

Recent studies in a limited gene set indicate that uH2A is mostly linked to transcriptional repression (Cao, Tsukada et al. 2005; Osley, Fleming et al. 2006; Stock, Giadrossi et al. 2007; Nakagawa, Kajitani et al. 2008). Consistent with these studies, our genome-wide analysis indicate that Bmi1-dependent uH2A exhibits a broad enrichment at the most repressed genes in the mouse genome (Fig. 4-5A,B). Peak centered analysis of promoter uH2A confirms this result (Fig. 4-5C) and reveals that this enrichment is most likely a consequence of over-representation of the H3K27me3 mark (Fig. 4-5C, D & E). These results indicate that unlike H3K27me3 (Barski, Cuddapah et al. 2007; Mikkelsen, Ku et al. 2007), uH2A by itself is not an accurate predictor of *de novo* expression levels and could serve to explain earlier studies which have reported the presence of uH2A at actively transcribed genes (Levinger and Varshavsky 1982; Barsoum and Varshavsky 1985). Instead,

our results suggest that uH2A plays a much more refined role in the control of gene expression. Analysis of expression changes in Bmi1 knockout MEFs revealed that, on average, genes marked by promoter peaks of uH2A are upregulated upon Bmi1 knockout. In addition, the level of enrichment of these promoter uH2A peaks, as indicated by sequence tag density, reveal a clear increase in average fold change when higher density peaks are compared with lower density peaks (Fig. 4-5F). Thus, our work extends gene specific studies and confirms the existence of a genome wide link between uH2A and gene silencing.

## EXPERIMENTAL PROCEDURES

*Cell Lines, Cell Culture, and Antibodies:* Mouse embryonic fibroblast cells were isolated from Bmi1 null and Bmi1 wild-type littermates and immortalized by expression of the TBX2.pBabePURO construct (Jacobs, Keblusek et al. 2000). Both cell lines were maintained in Dulbecco's modified Eagles Medium supplemented with 10% (v/v) fetal bovine serum and 1X Penicillin/Streptomycin. Antibodies employed in this study are as follows: α-uH2A (Millipore, 05-678), α-Bmi1 (Millipore, 05-637), and α-H3(Abcam, 1791).

*Chromatin Immunoprecipitation Assays:* ChIP assays were performed as previously described (Cao, Tsukada et al. 2005) with the following alterations. Chromatin was prepared from one 15 $cm^2$ plate grown to 95% confluence. After nuclei isolation, the pellet was resusupended in Solution B (20mM Hepes pH 7.9, 25% [v/v] glycerol, 0.5% [v/v] NP-40, 0.42M NaCl, 1.5mM $MgCl_2$, 1mM CaCl), subjected to sonication, and then treated with 30 units of micrococcal nuclease for 15 minutes to ensure mononucleosome resolution. Immunoprecipitations were carried out using 15 μg of antibody. For qPCR detection, the percent of IP enrichment as compared to input was calculated for both WT and Bmi1 null ChIPs using SYBR GreenER (Invitrogen) and data is presented as the fold change in percent input of WT versus Bmi1 null cells. All detection primers are listed in Table 4-2.

*Solexa sequencing:* Immunoprecipitated DNA fragments were blunt-ended, ligated to Solexa adaptors and sequenced using the Illumina 1G Genome Analyzer as previously described(Barski, Cuddapah et al. 2007).

*Sequence mapping:* The 25 bp sequenced reads were obtained and mapped to the mouse genome (*mm*8 assembly) using the Solexa Analysis Pipeline, as previously described (Barski, Cuddapah et al. 2007). This yielded a BED file containing a total of 6.1 and 8.0 million

unique tags for Bmi1 wild-type, and Bmi1 null libraries, respectively. Summary files were then created for each library by counting the number of tags falling into 200 bp genomic bins. A Bmi1-dependent uH2A summary file was created by applying a scaling factor to the WT file to equalize total tag count and subtracting the Bmi1 null tag count from the WT tag count within each genomic bin. The raw and processed sequencing and microarray data has been deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE15909.

*Expression microarray experiments:* 7 μg of total RNA from both Bmi1 wild-type and null MEFs was used to synthesize cDNA. A custom cDNA kit from Life Technologies was used with a T7-$(dT)_{24}$ primer for this reaction. Biotinylated cRNA was then generated from the cDNA reaction using the BioArray High Yield RNA Transcript Kit. The cRNA was then fragmented in fragmentation buffer (5X fragmentation buffer: 200mM Tris-acetate, pH8.1, 500mM KOAc, 150mM MgOAc) at 94°C for 35 minutes before the chip hybridization. 15 μg of fragmented cRNA was then added to a hybridization cocktail (0.05 μg $\mu l^{-1}$ fragmented cRNA, 50 pM control oligonucleotide B2, *BioB*, *BioC*, *BioD*, and *cre* hybridization controls, 0.1 mg $ml^{-1}$ herring sperm DNA, 0.5 mg $ml^{-1}$ acetylated BSA, 100mM MES, 1M $[Na^+]$, 20mM EDTA, 0.01% [v/v] Tween 20). 10 μg of cRNA was used for hybridization. Affymetrix Mouse Genome 430 2.0 Arrays were hybridized for 16 hours at 45°C in the GeneChip Hybridization Oven 640. The arrays were washed and stained with R-phycoerythrin streptavidin in the GeneChip Fluidics Station 400. After this, the arrays were scanned with the Hewlett Packard GeneArray Scanner. Affymetrix GeneChip Microarray Suite 5.0 software was used for washing, scanning, and basic analysis. Sample quality was assessed by examination of 3' to 5' intensity ratios of certain genes.

*Cytosine Extension:* The assay was adapted from a previous report (Fujiwara and Ito 2002). Briefly, 100 ng of genomic DNA was digested to completion using *MspI* and *HpaII* (Fermentas) and the result was subjected to single nucleotide extension in the presence of biotin-labeled dCTP (Invitrogen). One-fiftieth of the final reaction was manually spotted on (+) nylon membrane, incubated in 0.4N NaOH, neutralized with 1X TBS, baked at 80˚C for 20 minutes, and blocked overnight at 65˚C in blocking buffer (4X SSPE, 6X Denhardt's, 300μg $ml^{-1}$ salmon sperm DNA, 0.1% [w/v] SDS). The blot was incubated with a 1:5000 dilution of streptavidin conjugated alkaline phosphatase (Pierce) in blocking buffer for 20 minutes at room temperature, washed 3 x 15 minutes in TBST, and visualized using BCIP/NBT solution (Sigma). The developed blot was scanned and signal was quantified using NIH ImageJ software (http://rsb.info.nih.gov/ij/).

*Determination of Bmi1-dependent peaks of uH2A:* Peak enrichment calls were generated by the TIROE program (algorithm described in (Ho, Jothi et al. 2009)). Briefly, input was set to the mapped tag BED file corresponding to wild-type ChIP-seq and with the Bmi1 null BED file set to background. Parameter cut-offs were set as follows: p-value(p) ≤ 1e-5, Fold enrichment cutoff (f) ≥ 5, and Average DNA fragment length(f) of 232.

*Gene analysis groupings:* REFSEQ gene coordinates were extracted from the UCSC table browser and the list was parsed to only consider one isoform per annotated gene. Isoform selection was carried out by first searching for the 3' most transcription start sites. If a single gene had more than one isoform sharing a single TSS, then the longest isoform was kept for analysis. Promoters were defined as -1kb to +1kb surrounding TSSs and previously published peak data sets (Mikkelsen, Ku et al. 2007; Meissner, Mikkelsen et al. 2008) composed of H3K4me3, H3K27me3, CpG nucleotide density calls, and HCP promoter

methylation analyses were re-mapped onto this gene list. High resolution mapping surrounding gene TSSs was accomplished by extracting per base pair Bmi1-dependent uH2A tag density reads at 200 bp intervals surrounding TSSs.

*Peak distribution analysis:* The location of TIROE uH2A peaks was carried out in relation to the alignable genome as previously defined (Mikkelsen, Ku et al. 2007).

*Standard Error calculation:* Whole genome averaging analysis error bars represent the standard error of the mean (s.e.m.) and *P* values were determined using student's *t* tests. Two-sided proportional testing, Wilcoxon signed-rank testing, and box plot generation were carried out using the R package.

**Figure 4-1.** Bmi1-dependent promoter uH2A enrichment correlates with promoter H3K27me3 enrichment. **(A)** Profiles of per base pair Bmi1-dependent uH2A enrichment across the TSS of well annotated genes. Error bars represent s.e.m. of per base pair average for all genes analyzed (*n*=15867). **(B)** Well-annotated genes were grouped by epigenetic modification state and the average normalized uH2A tag density was determined for each group. No K4/No K27 (*n*=4690), purple; both H3K4me3 and H3K27me3 (Bivalent) (*n*=1377), green; H3K27me3 only (*n*=888), blue; H3K4me3 only (*n*=8912), red; all genes (*n*=15867), grey.

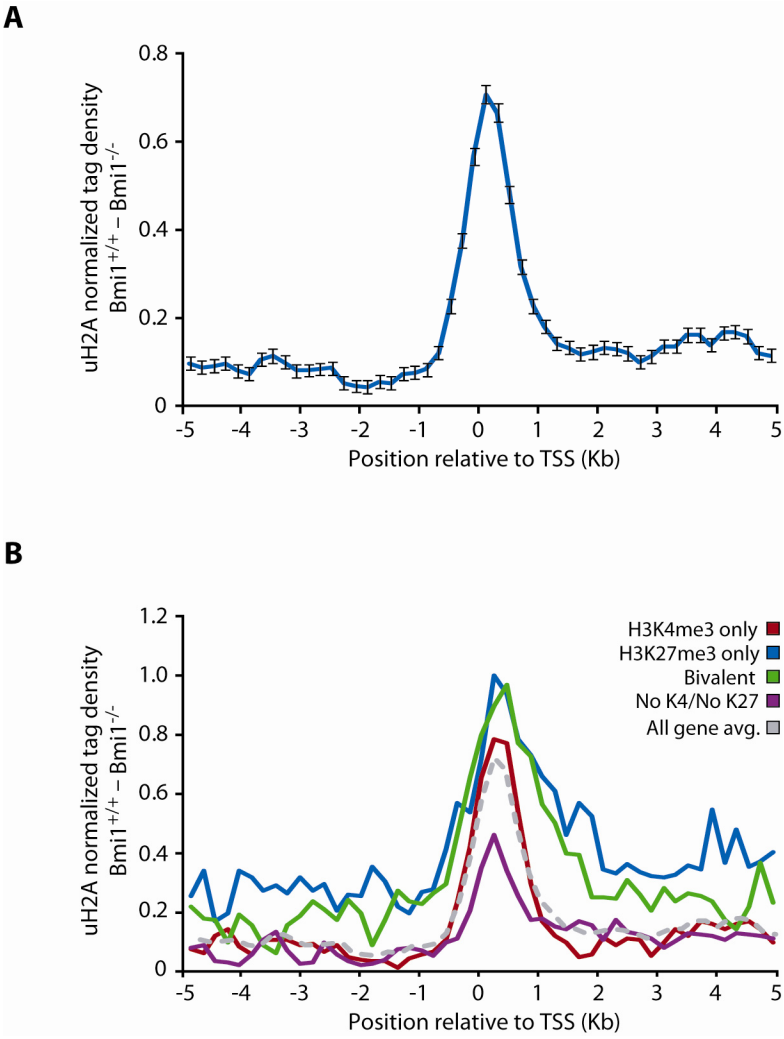

**Figure 4-2.** Bmi1-dependent promoter uH2A enrichment at a sub-set of gene promoters. **(A)** Bmi1-dependent tag density at selected gene promoters that contain peaks defined by the TIROE program. Red arrow indicates the start and direction of transcription. **(B)** ChIP-qPCR validation of uH2A peaks (blue bars) and Bmi1 enrichment (red bars) was carried out in WT and Bmi1 null MEF cells. Enrichment was normalized to input control and data is presented as $\log_2$ value of WT versus Bmi1 null.

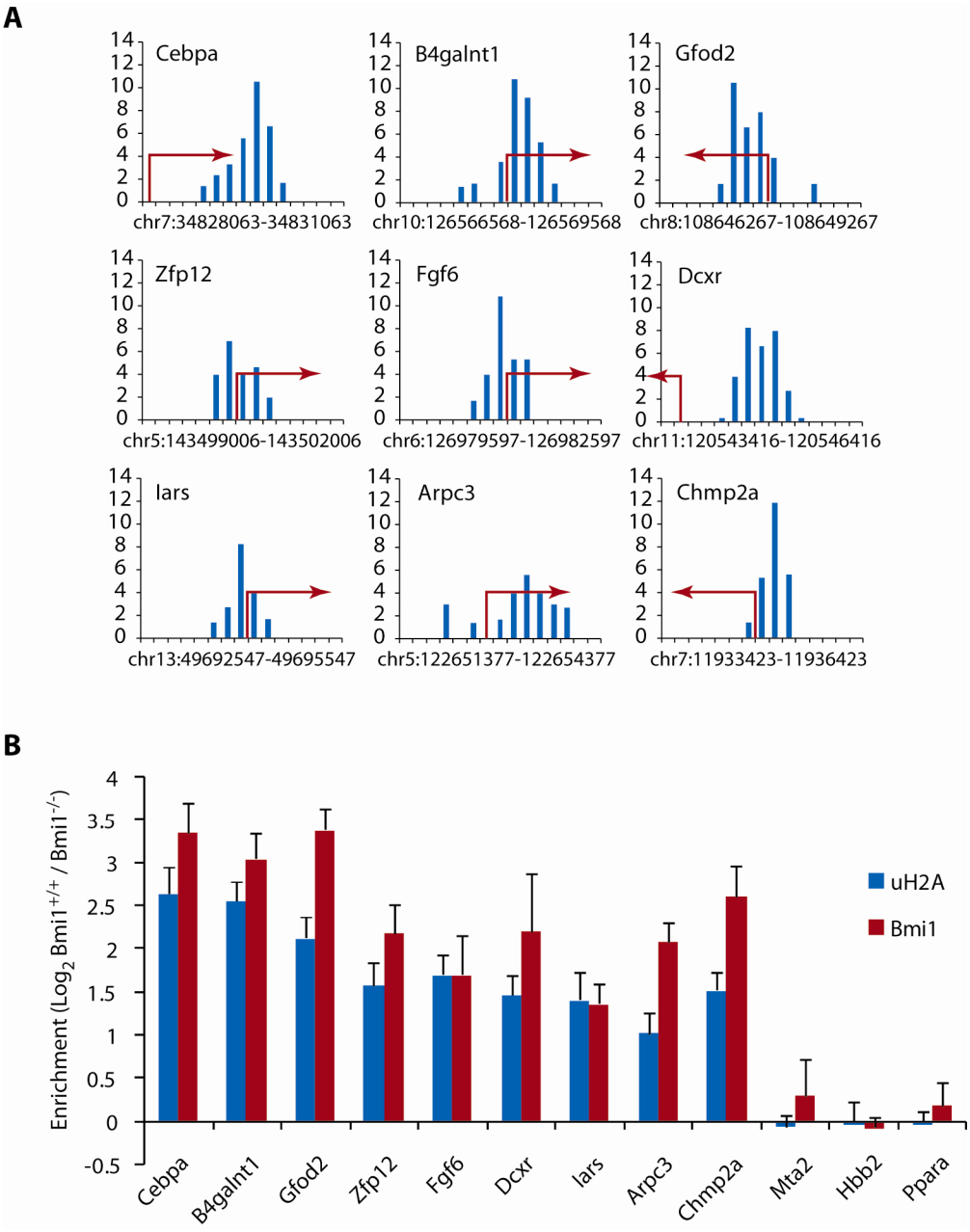

**Figure 4-3.** Promoter and non-promoter peaks of Bmi1-dependent uH2A are distinct. **(A)** Proportion of promoter uH2A peaks (*n*=671) (left bar) and non-promoter uH2A peaks (*n*=15735)(right bar) which overlap with H3K4/H3K27 methylation peaks. No K4/No K27, purple; both H3K4me3 and H3K27me3 (Bivalent), green; H3K27me3 only, blue; H3K4me3 only, red. **(B)** Box plot representation of peak tag density distribution for promoter and non-promoter uH2A peaks. Red lines indicate median values. *P* value derived from Wilcoxon signed-rank test. **(C)** Box plot representation of peak tag density distribution for promoter (+) and non-promoter (-) uH2A peaks further sub-divided by H3K4me3/H3K27me3 co-localization. No K4/No K27, purple; both H3K4me3 and H3K27me3 (Bivalent), green; H3K27me3 only, blue; H3K4me3 only, red. White lines indicate median values. *P* value derived from Wilcoxon signed-rank test.

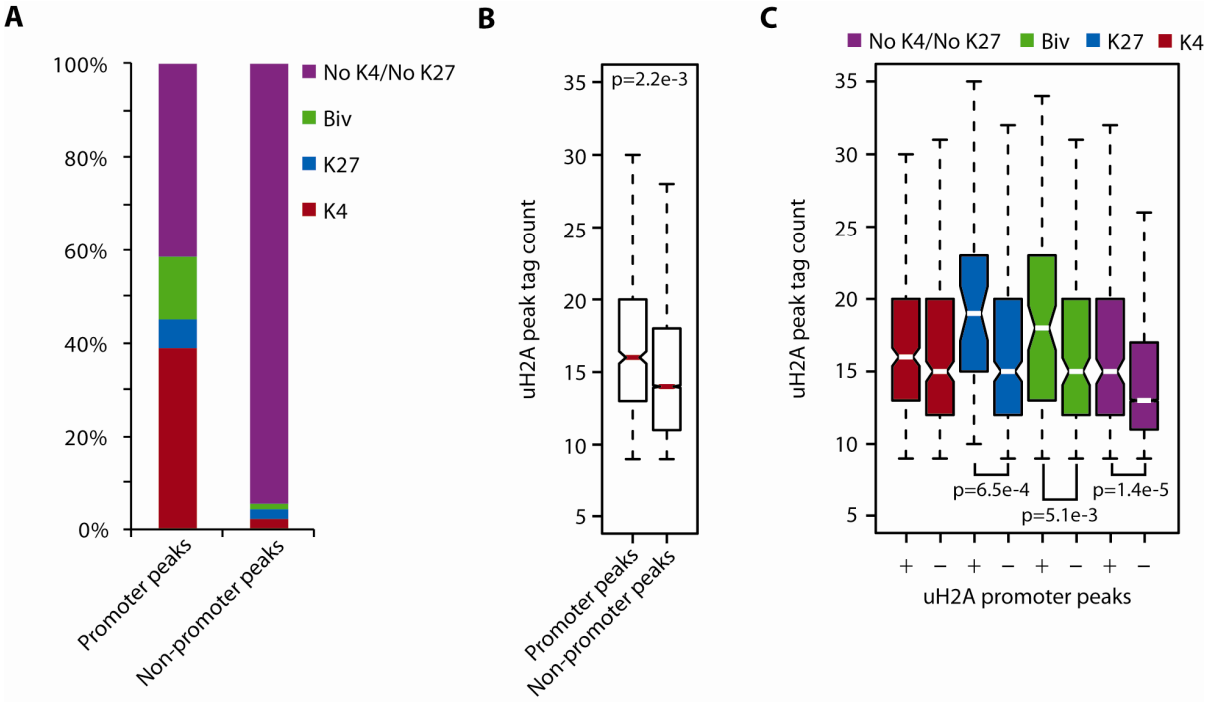

**Figure 4-4.** uH2A deposition is linked to DNA methylation. **(A)** Bmi1-dependent per base pair uH2A tag density was determined for genes defined as high CpG dinucleotide promoter content (HCP, blue) (*n*=10,310), intermediate CpG dinucleotide promoter content (ICP, red) (*n*=2889), low CpG dinucleotide promoter content (LCP, green) (*n*=2668). **(B)** HCP class genes with available DNA methylation data were grouped based on the presence (+) (*n*=266) or absence (-) (*n*=8230) of a promoter bound peak of Bmi1-dependent uH2A and the distribution of DNA methylation values for each group was visualized using a standard box plot. Red lines indicate median values. *P* value derived from Wilcoxon signed-rank test. **(C)** Western blotting of uH2A was carried out using cell lysate prepared from wild-type and Dnmt1 null MEFs. H3 was used as a loading control. **(D)** Cytosine extension analysis was carried out using wild-type and Bmi1 null MEFs as well as Dnmt1 null and control cells. Biotin labeled restriction digests were spotted in triplicate onto a Nylon membrane and visualized using alkaline phosphatase. The procedure was repeated three independent times and data for one representative experiment is shown. **(E)** Data presented in panel D were quantified and presented as the ratio between methylation sensitive *HpaII* incorporation versus methylation insensitive *MspI* incorporation. Error bars represent s.e.m. (*n*=3).

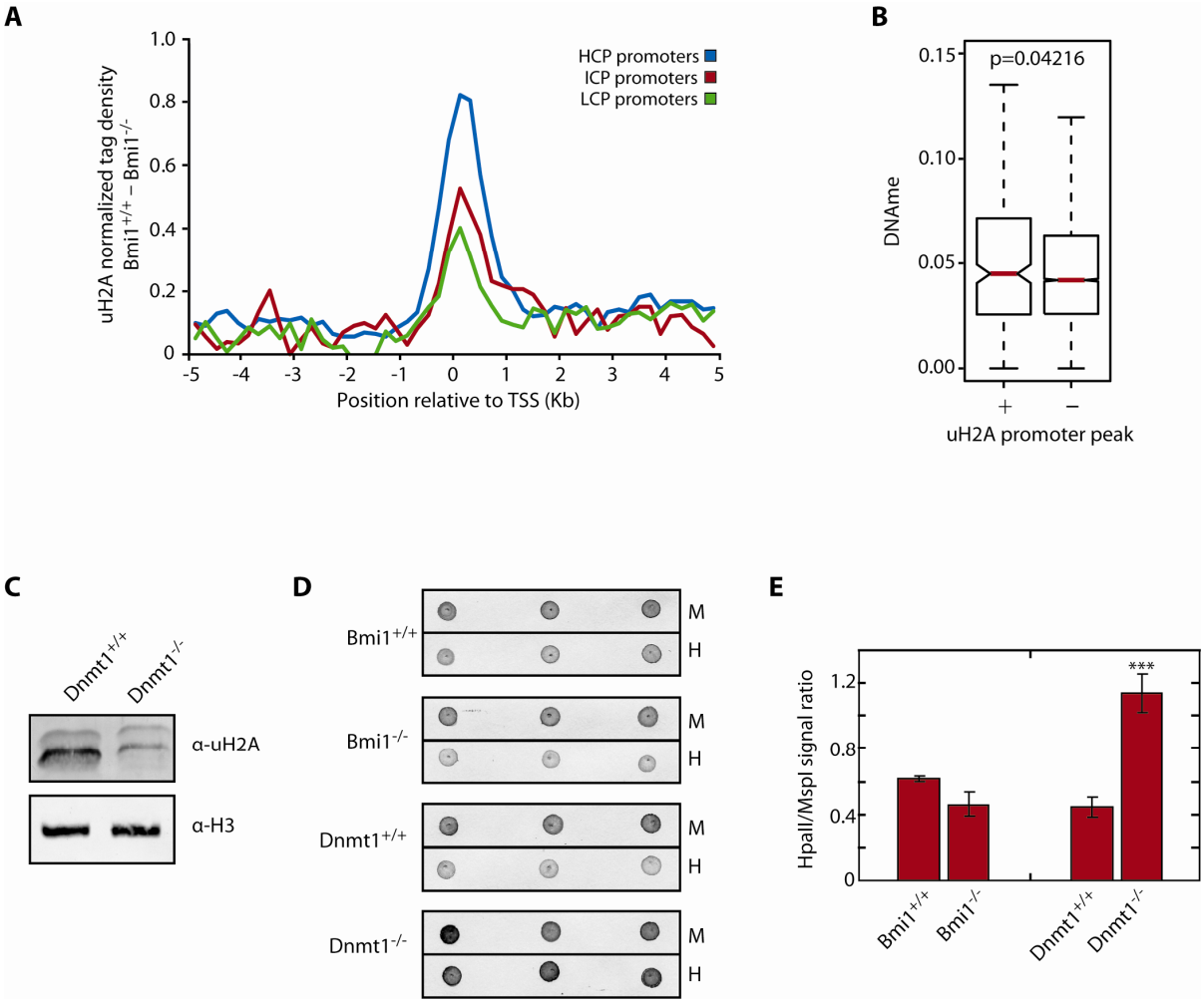

**Figure 4-5.** Bmi1-dependent uH2A is enriched at repressed genes. **(A)** Bmi1-dependent uH2A tag density surrounding the TSSs of genes was calculated and grouped into 10 bins ($n$≈1335 for each group) based on *de novo* gene expression levels in MEF cells. **(B)** Bmi1-dependent uH2A tag density was calculated for the region indicated in (A) for genes grouped based on *de novo* expression in MEF cells ($n$=4007, $n$=2674, $n$=2668, $n$=4005; highest expressed to lowest expressed). Error bars represent the s.e.m. of the highest expressed and lowest expressed gene groups. **(C)** WT MEF expression of genes with (+, $n$=483) and without (-, $n$=12,871) a peak of Bmi1-dependent uH2A was subjected to data distribution analysis by standard box plot. Red lines indicate median values. *P* value derived from Wilcoxon signed-rank test is indicated. **(D)** Gene promoters with (+) and without (-) peaks of uH2A were analyzed for the presence of H3K4/H3K27 methylation peaks. No K4/No K27, purple; both H3K4me3 and H3K27me3 (Bivalent), green; H3K27me3 only, blue; H3K4me3 only, red. **(E)** Box plot of expression data distribution for promoters with (+) and without (-) uH2A peaks in WT MEFs were further sub-divided by H3K4/H3K27 co-localization. No K4/No K27, purple; both H3K4me3 and H3K27me3 (Bivalent), green; H3K27me3 only, blue; H3K4me3 only, red. White lines indicate median values. *P* value derived from Wilcoxon signed-rank test is shown. **(F)** Genes containing a peak of Bmi1-dependent uH2A were divided into three groups based on increasing peak tag density ($n$=157, 157, and 158, respectively) and the distribution of $\log_2$ expression change (Bmi1 null versus WT MEF cells) was presented by box plot. Red lines indicate median values. *P* value derived from Wilcoxon signed-rank test is shown.

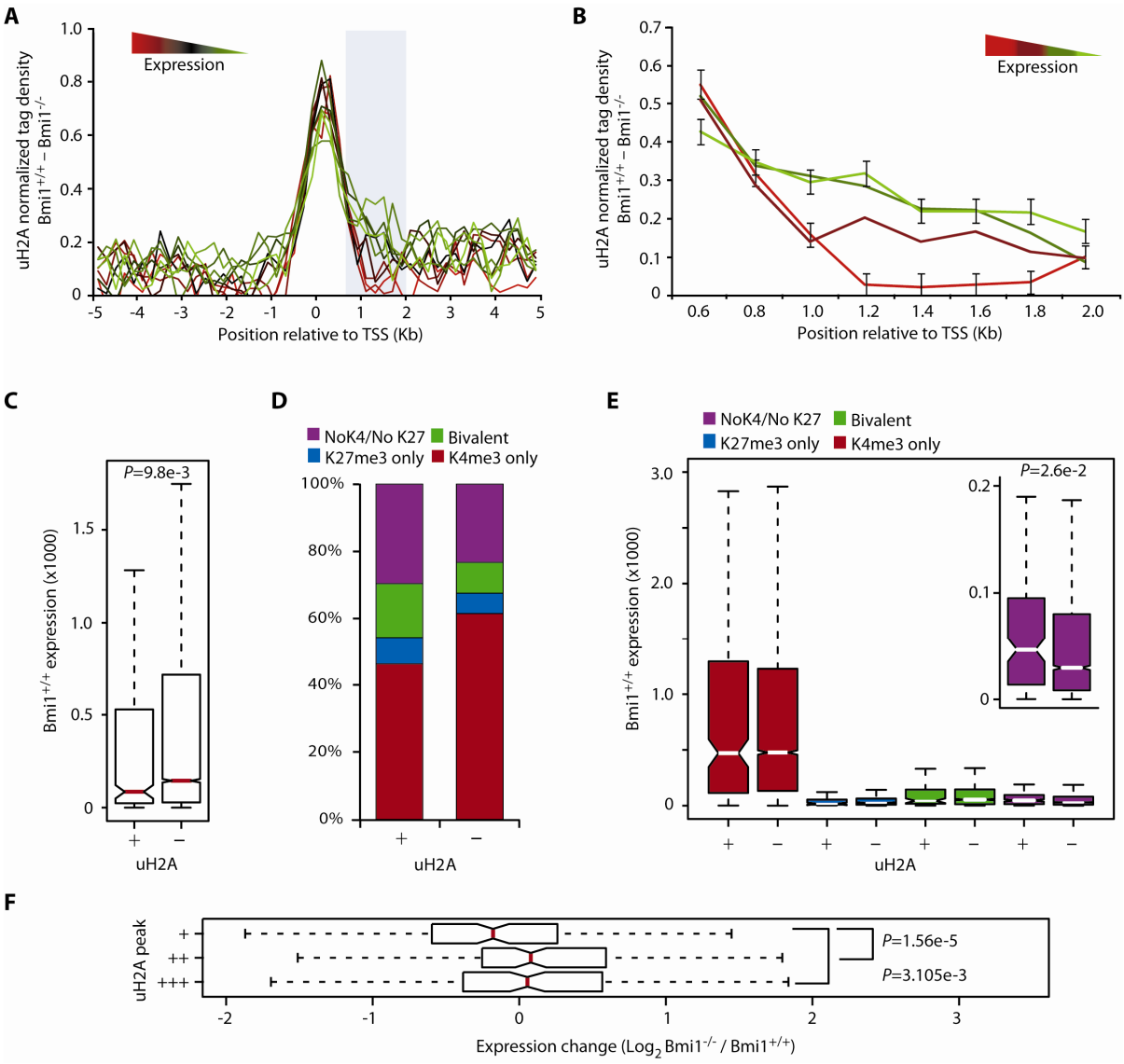
A
uH2A normalized tag density
Bmi1+/+ – Bmi1-/-
Expression
Position relative to TSS (Kb)
B
uH2A normalized tag density
Bmi1+/+ – Bmi1-/-
Expression
Position relative to TSS (Kb)
C
P=9.8e-3
Bmi1+/+ expression (x1000)
uH2A
D
NoK4/No K27
Bivalent
K27me3 only
K4me3 only
uH2A
E
NoK4/No K27
Bivalent
K27me3 only
K4me3 only
P=2.6e-2
Bmi1+/+ expression (x1000)
uH2A
F
uH2A peak
P=1.56e-5
P=3.105e-3
Expression change (Log2 Bmi1-/- / Bmi1+/+)

**Figure S4-1.** Genic peak distribution analysis reveals peak number enrichment towards the transcription termination site of genes and peak tag density enrichment within gene promoters. **(A)** Distribution histogram of peak location along the transcribed region of well-annotated genes (TSS, transcription start site; TES, transcription end site). **(B)** Genic uH2A peaks were grouped by their location within transcribed genes and the tag density data distribution was visualized by standard box plot. Red lines indicate median values. *P* value derived from Wilcoxon signed-rank test. *, **, and *** respectively indicate *P* value of 2.2e-14, 1.1e-15, 2.6e-10.

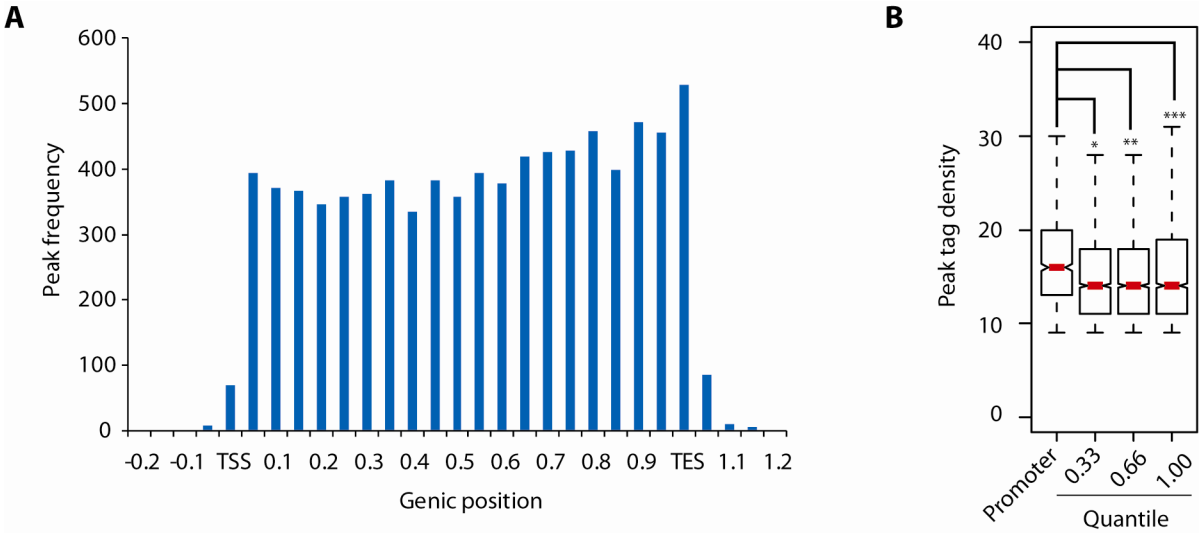

**Table 4-1.** Bmi1-dependent peak distribution throughout the mouse genome.

| | Genome* | Peak | | |
|---|---|---|---|---|
| | **BP coverage** | **Number** | **BP coverage** | ***P* value**** |
| **Gene & Promoter** | 893514855 (47.4%) | 8086 (49.3%) | 7748448 (52.4%) | <2.2e-16 |
| **Promoter** | 39206000 (2.1%) | 671 (4.1%) | 755102 (5.1%) | <2.2e-16 |
| **Transcribed Region** | 854308855 (45.3%) | 7415 (45.2%) | 6993346 (47.3%) | <2.2e-16 |
| **Non-genic** | 1770940233 (52.6%) | 8320 (50.7%) | 7046343 (47.6%) | <2.2e-16 |
| **Total** | 1884453825 (100%) | 16406 (100%) | 14794791 (100%) | |

***** Total base pairs refer to alignable portion of the genome.

****** Two-sided proportional test of Genome BP coverage as compared to Peak BP coverage.

**Table 4-2.** ChIP-qPCR primers used in the experiments described in Chapter 4.

| Gene | Forward Primer | Reverse Primer |
|---|---|---|
| Cebpa | CAGGGCAGGAGGAAGATACA | CACCTAAGTCCCTCCCCTCT |
| B4galnt1 | GGCGGATTTACGATCCAGT | GACTCCGGGGCTTTGTAGAC |
| Gfod2 | GGATGGGTTAGGTTCAAGCA | ACGGCATTGAGTACGAGGTC |
| Zfp12 | GTGCGCACTTCTGTTTGTGT | AAAGCTGCGAGCGTAGAGAC |
| Fgf6 | CTGAAGCAGGCTTTGGTTTC | ACCGCCCTTCTTGTTTTTCT |
| Dcxr | GAGCGCATACTCCTCCACTC | TGTTGGACTTGAGGTGGTCA |
| Iars | GGAGCGTCTTCTCCTTCCTT | TGCCATCCAACACCTACAAA |
| Arpc3 | ATTTCCCCCTTGTCATTTCC | TCCATCAGGGAAGTGAGGTC |
| Chmp2a | AACCAAATTAGGCCCACACA | GGTGCAAGCAATGGAAGAAT |
| Mta2 | TTCGTCTGGAGCTGAGCTTT | TCCCACCCGGTACATGTTAG |
| Hbb2 | GCCCAGGCTTAAGACATTTGAG | GAATTAGCTGCAAGGATAAGAACAGA |
| Ppara | GAGACCCTGGAGATGGTTGA | GTGTGCTTGTGTGCATGTGA |

## REFERENCES


Barrera, L. O. and B. Ren (2006). The transcriptional regulatory code of eukaryotic cells--insights from genome-wide analysis of chromatin organization and transcription factor binding. *Curr Opin Cell Biol* **18**(3): 291-8.

Barski, A., S. Cuddapah, et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* **129**(4): 823-37.

Barsoum, J. and A. Varshavsky (1985). Preferential localization of variant nucleosomes near the 5'-end of the mouse dihydrofolate reductase gene. *J Biol Chem* **260**(12): 7688-97.

Boyer, L. A., K. Plath, et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**(7091): 349-53.

Bracken, A. P., N. Dietrich, et al. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* **20**(9): 1123-36.

Buchwald, G., P. van der Stoop, et al. (2006). Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. *Embo J* **25**(11): 2465-74.

Cao, R., Y. Tsukada, et al. (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol Cell* **20**(6): 845-54.

Cao, R., L. Wang, et al. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**(5595): 1039-43.

Cao, R. and Y. Zhang (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr Opin Genet Dev* **14**(2): 155-64.

Chen, X., H. Xu, et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**(6): 1106-17.

de Napoles, M., J. E. Mermoud, et al. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev Cell* **7**(5): 663-76.

Elderkin, S., G. N. Maertens, et al. (2007). A phosphorylated form of Mel-18 targets the Ring1B histone H2A ubiquitin ligase to chromatin. *Mol Cell* **28**(1): 107-20.

Farthing, C. R., G. Ficz, et al. (2008). Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet* **4**(6): e1000116.

Fischle, W., Y. Wang, et al. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* **17**(15): 1870-81.

Fouse, S. D., Y. Shen, et al. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**(2): 160-9.

Fujiwara, H. and M. Ito (2002). Nonisotopic cytosine extension assay: a highly sensitive method to evaluate CpG island methylation in the whole genome. *Anal Biochem* **307**(2): 386-9.

Hernandez-Munoz, I., P. Taghavi, et al. (2005). Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1. *Mol Cell Biol* **25**(24): 11047-58.

Ho, L., R. Jothi, et al. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Natl Acad Sci U S A*.

Jackson-Grusby, L., C. Beard, et al. (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat Genet* **27**(1): 31-9.

Jacobs, J. J., P. Keblusek, et al. (2000). Senescence bypass screen identifies TBX2, which represses Cdkn2a (p19(ARF)) and is amplified in a subset of human breast cancers. *Nat Genet* **26**(3): 291-9.

Jacobs, J. J., K. Kieboom, et al. (1999). The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. *Nature* **397**(6715): 164-8.

Jenuwein, T. and C. D. Allis (2001). Translating the histone code. *Science* **293**(5532): 1074-80.

Kim, S. Y., S. W. Paylor, et al. (2006). Juxtaposed Polycomb complexes co-regulate vertebral identity. *Development* **133**(24): 4957-68.

Kim, T. H., Z. K. Abdullaev, et al. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**(6): 1231-45.

Kondo, Y., L. Shen, et al. (2008). Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation. *Nat Genet* **40**(6): 741-50.

Kotake, Y., R. Cao, et al. (2007). pRB family proteins are required for H3K27 trimethylation and Polycomb repression complexes binding to and silencing p16INK4alpha tumor suppressor gene. *Genes Dev* **21**(1): 49-54.

Ku, M., R. P. Koche, et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**(10): e1000242.

Lagarou, A., A. Mohd-Sarip, et al. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev* **22**(20): 2799-810.

Lee, M. G., R. Villa, et al. (2007). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* **318**(5849): 447-50.

Lee, T. I., R. G. Jenner, et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**(2): 301-13.

Levinger, L. and A. Varshavsky (1980). High-resolution fractionation of nucleosomes: minor particles, "whiskers," and separation of mononucleosomes containing and lacking A24 semihistone. *Proc Natl Acad Sci U S A* **77**(6): 3244-8.

Levinger, L. and A. Varshavsky (1982). Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the Drosophila genome. *Cell* **28**(2): 375-85.

Lewis, A., K. Mitsuya, et al. (2004). Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nat Genet* **36**(12): 1291-5.

Martin, C. and Y. Zhang (2005). The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* **6**(11): 838-49.

Meissner, A., T. S. Mikkelsen, et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*.

Mikkelsen, T. S., M. Ku, et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**(7153): 553-60.

Min, J., Y. Zhang, et al. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* **17**(15): 1823-8.

Nakagawa, T., T. Kajitani, et al. (2008). Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. *Genes Dev* **22**(1): 37-49.

Negishi, M., A. Saraya, et al. (2007). Bmi1 cooperates with Dnmt1-associated protein 1 in gene silencing. *Biochem Biophys Res Commun* **353**(4): 992-8.

Olson, M. O., I. L. Goldknopf, et al. (1976). The NH2- and COOH-terminal amino acid sequence of nuclear protein A24. *J Biol Chem* **251**(19): 5901-3.

Osley, M. A., A. B. Fleming, et al. (2006). Histone ubiquitylation and the regulation of transcription. *Results Probl Cell Differ* **41**: 47-75.

Sakamoto, Y., S. Watanabe, et al. (2007). Overlapping roles of the methylated DNA-binding protein MBD1 and polycomb group proteins in transcriptional repression of HOXA genes and heterochromatin foci formation. *J Biol Chem* **282**(22): 16391-400.

Schones, D. E. and K. Zhao (2008). Genome-wide approaches to studying chromatin modifications. *Nat Rev Genet* **9**(3): 179-91.

Stock, J. K., S. Giadrossi, et al. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* **9**(12): 1428-35.

Umlauf, D., Y. Goto, et al. (2004). Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nat Genet* **36**(12): 1296-300.

Vassilev, A. P., H. H. Rasmussen, et al. (1995). The levels of ubiquitinated histone H2A are highly upregulated in transformed human cells: partial colocalization of uH2A clusters and PCNA/cyclin foci in a fraction of cells in S-phase. *J Cell Sci* **108 (Pt 3)**: 1205-15.

Vire, E., C. Brenner, et al. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**(7078): 871-4.

Wang, H., L. Wang, et al. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**(7010): 873-8.

Wang, L., J. L. Brown, et al. (2004). Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* **14**(5): 637-46.

Wang, Z., C. Zang, et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**(7): 897-903.

Wu, X., Y. Gong, et al. (2008). Cooperation between EZH2, NSPc1-mediated histone H2A ubiquitination and Dnmt1 in HOX gene silencing. *Nucleic Acids Res* **36**(11): 3590-9.

Xu, X., M. Bieda, et al. (2007). A comprehensive ChIP-chip analysis of E2F1, E2F4, and E2F6 in normal and tumor cells reveals interchangeable roles of E2F family members. *Genome Res* **17**(11): 1550-61.

Zhou, W., P. Zhu, et al. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Mol Cell* **29**(1): 69-80.