2009

Regulated Histone H3 Proteolysis During Mouse Embryonic Stem Cell Differentiation

Elizabeth M. Duncan

REGULATED HISTONE H3 PROTEOLYSIS DURING MOUSE EMBRYONIC STEM CELL DIFFERENTIAION

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Elizabeth M. Duncan

June 2009

REGULATED HISTONE H3 .PROTEOLYSIS DURING MOUSE EMBRYONIC STEM CELL DIFFERENTIATION

Elizabeth M. Duncan, Ph.D.

The Rockefeller University 2009

The association of genomic DNA with histone proteins in the three-dimensional structure known as chromatin is the central framework for "epigenetics," which is defined as inherited phenotypes governed by differences that cannot be explained by changes in DNA sequence. In recent years, studies have shown that regulated changes in the chemical and physical properties of chromatin often lead to dynamic changes in many cellular processes, including development and differentiation, by affecting the accessibility of the genomic information stored in the DNA.

The cell uses many different mechanisms to regulate chromatin in order to establish, maintain, and propagate patterns of gene expression that are necessary for proper development and differentiation. Many of these mechanisms involve the histone component of chromatin, both through chemical and structural changes of the histone proteins themselves and via complex interactions with other non-histone chromatin proteins. Here, in my thesis work, I describe how a few of these chromatin regulatory mechanisms are used during mammalian differentiation, specifically focusing on those involving histone H3.

First, in Chapter 2, I describe how certain non-histone chromatin proteins that are key to development specifically interact with modified histones using biochemical, biophysical and structural approaches. Next, in Chapter 3, I describe how chromatin undergoes specific, dramatic changes as cells lose their capacity for

self-renewal and proceed toward a specific lineage using a mouse embryonic stem cell model of differentiation and early embryonic development. These changes involve both the post-translational modification of histone H3 and the incorporation of different H3 variant proteins into the chromatin fiber. Finally, I describe the observation that differentiating mouse embryonic stem cells proteolytically cleave histone H3 and identify a protease that accomplishes this cleavage.

ACKNOWLEDGMENTS

First, I would like to thank my adviser and mentor, Dr. David Allis, for his support and guidance. Although I was at first afraid to show him my “mysterious” histone H3 blots, particularly after ignoring most of his suggestions for a thesis project, Dave enthusiastically supported my quest to discover their identify – sometimes more than I did myself. His optimism and excitement were often a much-needed remedy for my self-doubt, and I greatly appreciate his constant support and encouragement. Dave’s enthusiasm also translated into many creative ideas, both his own and those he sparked in me.

I thank my faculty advisory committee, Dr. Sanford Simon, Dr. Thomas Muir, Dr. Robert Roeder and Dr. Shai Shaham for their time, their thoughtfulness about my work, and all their helpful suggestions. In particular, I thank Dr. Simon for serving as the chair of my committee and Dr. Shahin Rafii for serving on my thesis committee as the external member.

I must also thank Dr. Jan Breslow and Dr. Ephraim Sehayek for their invaluable guidance and support when I was a research assistant in the Breslow Laboratory. They not only bravely gave me the opportunity to learn science, but also set me on the path to becoming a scientist.

There are many past and present members of the Allis Laboratory to whom I owe many thanks. Dr. Emily Bernstein served not only as my “lab coach” when I first started working in the Allis laboratory, but also as an excellent role model of a successful scientist and conscientious lab colleague. I must also thank Dr. Sandra Hake and Dr. Joanna Wysocka, my lab “sister”, for all the time and attention they spent answering my never-ending questions and guiding me both technically and intellectually.

I thank all my Allis lab colleagues for their support and insightful discussions. In particular, I thank Aaron Goldberg for helping to pioneer the embryonic stem cell model in the lab, for collaborating with me on the ChIP-seq experiments and for his critical reading of part of my thesis. I also thank both Robert Diaz and Jamie Winshell for their enormous support and conscientiousness in running our laboratory. In addition, I thank Lindsey Baker, Laura Banaszynski, Ping Chi and Jamie Winshell for all their time and advice on my thesis presentation.

I am hugely indebted to our collaborators in the Hunt Laboratory, Dr. Tara Muratore-Schroeder, Dr. Benjamin Garcia and Dr. Donald Hunt for their meticulous mass spectrometry work and many thoughtful science discussions. My thesis work would be much lesser without their contribution. I also thank our collaborators in the Patel Laboratory, particularly Dinshaw Patel and Haitao Li, with whom I was lucky to work on several projects and from whom I learned many things.

I thank the scientists at the Proteomics Resource Center, who have been very helpful with peptide and protein projects, and Scott Dewel at the Genomics Resource Center, who generated all the Solexa sequencing data and patiently taught Aaron and I how to analyze it. I also thank Deyou Zheng at Albert Einstein College of Medicine for all his analysis of the data, far beyond what I am able to do.

Many thanks to the members of David Rockefeller Graduate Program Dean's office for their help both over the years and, most intensely, these last few weeks.

I thank the Tri-Sci Stem Cell Institute, the Anderson Cancer Center and the David Rockefeller Graduate Program their vital financial support.

Finally, I thank my family. My parents and my husband, in particular, have provided constant support and encouragement throughout this long journey. I also thank my daughter, who kindly slept through (most) nights while I was writing this report.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

aa	amino acid
Ab	antibody
AEBSF	4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride
bp	base pairs
BPTF	Bromodomain-PHD-finger transcription factor
BSA	bovine serum albumin
C	celsius
CathL	Cathepsin L
CBX	chromobox
CD	chromo domain
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
Chromo	Chromatin organization modifier
2D	two dimensional
DAPI	4',6-diamidino-2-phenylindole
DAVID	Database for Annotation, Visualization and Integrated Discovery
DMEM	Dulbecco's Modified Eagle's Medium
DNA	deoxyribonucleic acid
ssRNA	single-stranded RNA
DTT	DL-1,4-dithiothreitol
E-64	trans-Epoxysuccinyl-L-leucylamido(4-guanidino)butane
EB	embryoid body
EDTA	ethylenediaminetetraacetic acid
EGFP	enhanced green fluorescent protein
ELISA	Enzyme-linked immunosorbent assay
ESC	embryonic stem cell
EZH2	enhancer of zeste homologue 2
FAM	Ferrocenyl substituted AziridinylMethanol
FACS	Fluorescence activated cell sorting
FL	full-length
FPLC	fast protein liquid chromatography
g	grams
gen	general

GFP	green fluorescent protein
GST	glutathione-S-transferase
h	hour
H3	histone H3
H3cs.1	histone H3 cleavage site 1
HAT	histone acetyltransferase
HDAC	histone deacetylase
HeLa	Henrietta Lacks
HEPES	2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulfonic acid
His	histidine
HMTase	histone methyltransferase
Hox	homeobox
HP1	Heterochromatin Protein 1
HPLC	high performance liquid chromatography
HRP	horse-radish peroxidase
IF	immunofluorescence
IP	immunoprecipitation
IPTG	isopropylthio-b-D-galactoside
K	lysine
kb	kilo bases
Kd	dissociation constant
kD	kilo dalton
KO	knock out
LB	Luria-Bertani (broth)
LIF	leukemic inhibitory factor
LTQ-FT	linear quadrupole ion trap-Fourier transform mass spectrometer
m	meter
M	molar
MBT	malignant brain tumor
MEF	mouse embryonic fibroblast
min	minute
mRNA	messenger RNA
MS	mass spectrometry
MW	molecular weight

N2a	neuro 2 a
Ni-NTA	nickel nitrilo-triacetic acid
NURF	Nucleosome remodeling factor
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
Pc	Polycomb
PcG	Polycomb group
PBS	phosphoate buffered saline
PCR	polymerase chain reaction
PEV	position effect variegation
PHD	plant homeodomain
PMSF	phenylmethanesulfonyl fluoride
PVDF	polyvinylidene fluoride
Q-PCR	quantitative polymerase chain reaction
r	recombinant
RA	retinoic acid
RD	replication dependent
REMSA	RNA electrophoretic mobility-shift assay
RNA	ribonucleic acid
RNase	ribonuclease
RNAi	RNA interference
RP-HPLC	reversed-phase high-performance liquid chromatography
RT	room temperature
RT-PCR	reverse transcriptase PCR
s	second
SD	standard deviation
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Su(var)	suppressor of variegation
TAU	triton acid urea
Tris	tris(hydroxymethyl)minomethane
trxG	thrithorax group
WT	wild-type, original sequence without mutation
X_i	inactive X chromosome

Chapter 1: General Introduction

A fundamental question in biology is how a single genome is able to dictate the differentiation and development of many different cell types. *In vivo*, the genome of a cell is packaged into a three-dimensional structure known as chromatin: DNA wraps around a core of histone proteins to form a chain of repeating units, or "beads on a string," that is then further compacted into more complex fibers (**Figure 1.1A**). As a result, the accessibility of the genome to transcription factors and other proteins is governed by the nature of this compaction (Van Holde, 1989). Chromatin that is less compact, and therefore appears lighter when stained with basic dyes, is termed euchromatin (**Figure 1.1B**); it is more accessible to the gene expression machinery of the cell, allowing those genes to be actively expressed (Van Holde, 1989). Chromatin that is highly compacted, and therefore appears dark by the same stains, is categorized as heterochromatin (**Figure 1.1B**); tighter compaction makes the genetic loci in heterochromatin less accessible, causing those genes to be silent (Van Holde, 1989). The association of genomic DNA with histone proteins thus provides the central framework for "epigenetics," which is defined as inherited phenotypes governed by differences that cannot be explained by changes in DNA sequence. It is therefore likely that both DNA and histone proteins participate in the mechanisms that regulate self-renewal and differentiation during cellular development.

Although for many years histone proteins were regarded as an inert scaffold for the genome, studies in our laboratory and others have shown that they are in fact dynamically regulated, both chemically and physically, in ways that lead to functional change. The cell uses various mechanisms to achieve this regulation, including post-translational modification of histone proteins, the remodeling of chromatin by ATP-dependent complexes (including nucleosome sliding and

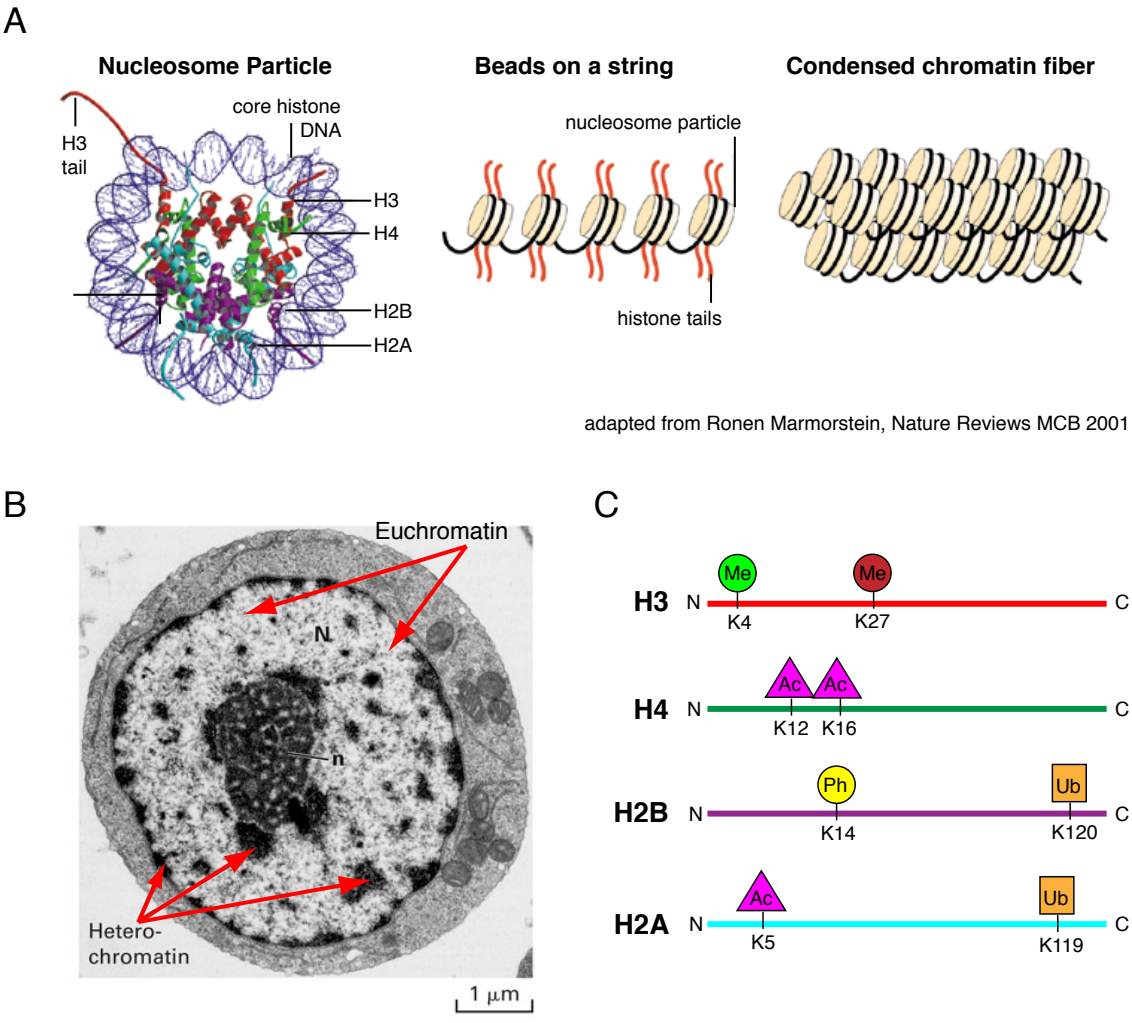


Figure 1.1: The organization of chromatin.

A. At the smallest level, the crystal structure of a nucleosome particle shows the histone octomer wrapped by the DNA helix (Luger et al., 1997); these particles are then linked together by the DNA to form the classic "beads on a string" organization; the string is then further condensed into a compact chromatin fiber in order to fit into the cell nucleus. (Images adapted from (Marmorstein, 2001).)

B. An electron micrograph of a mammalian cell nucleus (N); red arrows indicate the dark heterochromatic regions and the light euchromatic regions; n = nucleolous. (Image adapted from (Lodish et al., 2008).)

C. The four core histone proteins are modified by many different covalent post-translational modifications. Shown here are a few well documented examples, including histone methylation (green and red circles) on H3K4 and H3K27; acetylation (triangles) on H2AK5, H4K12, and H4K16; phosphorylation (yellow circles) at H2BK14, and ubiquitination (orange boxes) at H2AK119 and H2BK120.

replacement), and the incorporation of variant histone isoforms into the nucleosome particle (Ahmad and Henikoff, 2002; Allfrey et al., 1964; Varga-Weisz et al., 1995). Since cells use these chromatin regulatory pathways to establish, maintain and propagate different patterns of gene expression during normal development and differentiation it then follows that errors made in their regulation of the "epigenome" could lead to inappropriate expression or silencing of genes and, in turn, alter cellular identity. Moreover these alterations could lead to pathological situations, such as tumorigenesis, in certain physiological and developmental contexts.

In our laboratory and others in the chromatin field, we are interested in uncovering the ways in which the structural and biochemical properties of chromatin impact the biology of the cell. In particular, we are interested in the many different ways the amino terminal tails of core histone proteins (H2A, H2B, H3 and H4) are chemically modified to create a dynamic functional readout (**Figure 1.1C**). These modifications include, but are not limited to, acetylation, methylation, phosphorylation and ubiquitination (Allis et al., 2006). The combination of multiple types of modifications and multiple, yet specific, acceptor sites for them in each histone tail allows the cell to create many different biochemical patterns in its chromatin and, therefore, to relay and/or store many different bits of information without requiring any change to the DNA sequence.

Here I describe my research in the Allis laboratory on the histone H3 protein, including studies on its biochemical profile, its interactions with other non-histone chromatin proteins, and its developmentally-regulated proteolysis. For much of this work I used a mouse embryonic stem cell (ESC) model in order to study changes in chromatin, and histone H3 in particular, during the process of cellular differentiation. Below I will briefly summarize the background theory and literature for each of these studies and the questions that they aim to address.

Post-translational Modifications of Histone Proteins

In addition to their role in packaging the DNA into the nucleus, histones were also proposed to inhibit DNA function, i.e. RNA synthesis, from as early as 1951 (Stedman and Stedman, 1951), although clear evidence for this effect was not presented for another decade (Allfrey et al., 1963; Allfrey and Mirsky, 1962; Huang and Bonner, 1962; Stedman and Stedman, 1951). Vincent Allfrey and colleagues showed that histones inhibit RNA synthesis in nuclei isolated from calf thymus (Allfrey and Mirsky, 1962) and that they could increase the rate of RNA synthesis by removing histones from the nucleus (Allfrey et al., 1963; Allfrey and Mirsky, 1962). In subsequent work, they then presented evidence that histones were acetylated after translation and that this modification could affect their ability to inhibit RNA synthesis, despite the fact that such acetylated histones still retained a high affinity for DNA (Allfrey et al., 1964). The Grunstein laboratory later demonstrated these effects in living cells, having pioneered the use of yeast genetics to study the functional effects of histone deletions and mutations (Durrin et al., 1991; Han and Grunstein, 1988; Han et al., 1988; Kayne et al., 1988; Mann and Grunstein, 1992). Importantly, Allfrey et al foresaw that the modification of histones might be part of a more subtle mechanism to regulate both the activation and repression of transcription at different genomic loci, rather than simply blocking all RNA synthesis (Allfrey et al., 1964).

When James Brownell and colleagues identified the transcription factor GCN5 as an enzyme responsible for acetylating histones in 1996, they not only made the first identification of a histone-modifying enzyme but also provided the first direct link between a histone modification and cellular function, namely transcription (Brownell et al., 1996). Strong support of this finding occurred nearly simultaneously when Taunton et al identified of a human enzyme with histone deacetlyase activity that is highly similar to the yeast transcriptional regulator

Rpd3p (Taunton et al., 1996). Since then many histone-modifying enzymes have been identified, including Histone Acetyl Transferases (HATs) like GCN5, Histone DeACetylases (HDACs) like Rpd3p, Histone Methyl Transferases (HMTases), and kinases that phosphorylate specific histone residues, many of which have also been shown to play roles in transcription (de Ruijter et al., 2003; DeSouza et al., 2000; Hsu et al., 2000; Rea et al., 2000; Roth et al., 2001).

Although histone acetylation had been correlated with gene transcription long before the identification of GCN5, the role of histone methylation was less clear until several responsible HMTases were identified and shown to associate with particular regions of chromatin and/or gene promoters (Kouzarides, 2002). For example, the first lysine methyltransferase identified, Suppressor of variegation 39 (Suvar39), was already known to be enriched at heterochromatin when Rea et al demonstrated its specificity for H3K9 (Rea et al., 2000). Moreover, position-effect-variegation (PEV) is the phenomenon by which an actively expressed gene is silenced as the result of a chromosomal rearrangement that positions it next to heterochromatin. Suppressors of this phenomenon restore normal gene expression when mutated, and therefore function to silence gene expression in wild-type cells (**Figure 1.2A**). Rea et al also mapped the catalytic activity of Suvar39 to its evolutionarily conserved SET domain, which is found in many chromatin-associated proteins.

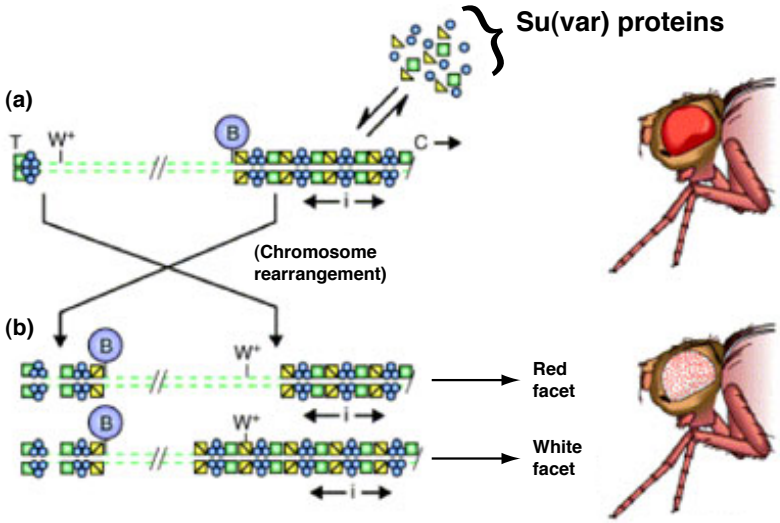
Two groups of proteins that both possess HMTase activity and play important roles in development are the trithorax group (trxG) and polycomb group (PcG) proteins. Their genes were discovered in *Drosophila* as activators and repressors of HOmeoboX (Hox) gene expression, respectively. Hox gene regulation is critical to the development of metazoans because they encode transcription factors that establish body patterning along the anterior posterior axis (Pearson et al., 2005). Additional studies using Chromatin ImmunoPrecipitation (ChIP) and DNA

Figure 1.2: The "active" or "silent" state of chromatin is determined by the actions of non-histone chromatin proteins.

A. A schematic of the position-effect-variegation (PEV) phenomenon: (a) the wild-type white gene (W+), which produces red-eyed flies when expressed, is normally located in a euchromatic region of chromatin and therefore not subjected to the silencing effects of heterochromatic Su(var) proteins since they do not spread beyond the hypothetical barrier element, B; (b) a chromosomal rearrangement (resulting from the repair of chromosomal breaks induced by irradiation) positions the white gene next to a heterochromatic region and moves the barrier element further along the chromosome, allowing the Su(var) proteins to spread beyond their normal limit in a stochastic fashion (i = initiation binding sites). This results in the silencing of the white gene in some cells of the eye, giving a mottled (or variegated) appearance. (Cartoon adapted from (Grewal and Elgin, 2002).)

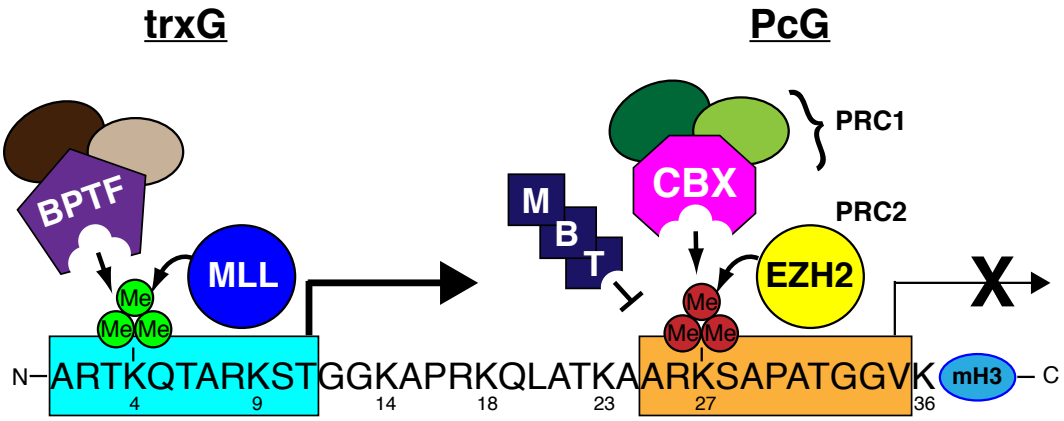
B. An illustration of the enzymatic machinery responsible for writing and reading histone methylation marks. The trithorax group (trxG) proteins include histone methyltransferases (HMTases) such as human MLL that methylate histone H3 lysine 4, which is correlated with gene activation (thick arrow); PHD finger proteins such as hBPTF, the largest subunit of the human NUcleosome Remodeling Factor (NURF) complex, are then able to bind, or "read", this mark and recruit additional complex members (brown ovals) in order initiate downstream remodeling events that lead to changes in gene expression. The polycomb group (PcG) proteins, on the other hand, are required for gene silencing and include HMTases such as the Polycomb Repressive Complex 2 (PRC2) protein hEZH2, which methylates H3K27. The chromodomain (CD)-containing PRC1 complex proteins (i.e. *Drosophila* Pc and mammalian CBX2, 4, 6, 7, and 8) can then engage this mark and recruit other PcG proteins (green ovals) that mediate transcriptional repression. PcG proteins that contain MBT repeats rather than CDs, such as L3MBT and SFMBT, bind mono and di, but not tri, methylated lysines.

A



adapted from Grewal and Elgin, Current Opinion in Genetics & Development 2002

B



Transcriptional activation

HMTases:
dmTrx & Ash1, vertebrate MLL (hMLL partially rescues fly Trx mutant)

Effectors:
PHD finger proteins (me3) including BPTF/NURF301, ING2

Transcriptional repression

HMTases:
dmE(z), mammalian EZH2

Effectors:
dmPc, mammalian Cbx2, 4, 6, 7, 8 (Cbx2 partially rescues fly Pc mutant; others not known)
MBT repeat proteins (me1,2) including L3MBT, SFMBT

Figure 1.2: The "active" or "silent" state of chromatin is determined by the actions of non-histone chromatin proteins.

sequencing have shown that trxG and PcG proteins target many other genes as well, including several families of key developmental regulators in addition to the Hox genes (Bracken et al., 2006; Lee et al., 2006; Milne et al., 2005; Orlando et al., 1998). Importantly, the trxG proteins include the HMTases such as dmTrx, dmAsh1 and vertebrate MLL, all of which methylate histone H3 at lysine 4. On the other hand, PcG proteins include the H3K27 HMTases dmE(z) and mammalian EZH2. In other words, trxG and PcG proteins have a ying-yang relationship, in which trxG proteins methylate H3K4 and maintain gene activation, and PcG proteins methylate H3K27 and maintain gene repression (**Figure 1.2B**). The identification of the trxG and PcG complexes as those responsible for H3K4me3 and H3K27me3, respectively, contributed enormously to our understanding of the link between these particular covalent modifications and gene expression, particularly since their mutants were already known to affect that of Hox genes. Other studies demonstrated that the trxG and PcG protein complexes are necessary to maintain rather than initiate gene expression (Schuettengruber et al., 2007). Nevertheless, many questions still remain as to the exact mechanisms of trxG and PcG recruitment and the regulation of gene expression.

Histone Binding Proteins as Mediators of Downstream Function

The discovery of these histone-modifiying enzymes provided insight into the mechanisms that establish the various patterns of histone modifications. However, many questions still remained as to how these marks were propagated and, importantly, translated into functional readouts. Many laboratories, including ours, hypothesized that some chromatin-associated proteins specifically recognize individual histone modifications, and/or combinations of modifications, in order to recruit these chromatin proteins to certain regions of the genome (Fischle et al., 2003a). Indeed, researchers have shown that histone acetylation increases the affinity of bromodomain proteins (Dhalluin et al., 1999; Hudson et al., 2000;

Jacobson et al., 2000; Owen et al., 2000), methylation engages proteins with motifs such as the chromodomain and the PHD finger (Bannister et al., 2001; Lachner et al., 2001; Shi et al., 2006; Wysocka et al., 2006), and phosphorylation recruits proteins such as 14-3-3 (Macdonald et al., 2005). These recognition proteins then interact with other proteins and protein complexes to propagate the histone modification itself and/or trigger additional activity. Below I will describe the identification of three key methyl-binding motifs, since they pertain directly to my thesis work (**Figure 1.2B**).

One significant contribution to the understanding of how certain non-histone proteins, or "effectors," interact with specific post-translational histone modifications to mediate downstream function was the identification of the plant homeodomain (PHD) finger as an H3K4me3 binding motif (Shi et al., 2006; Wysocka et al., 2006). At the time, no other H3K4me binding motifs had been identified. This motif was identified simultaneously within the protein BPTF, the largest subunit of the human NUcleosome Remodeling Factor (NURF) complex (Wysocka et al., 2006) and the INhibitor of Growth, family member 2 (ING2) protein (Shi et al., 2006). The NURF chromatin-remodeling complex was originally purified from *Drosophila* embryos (Tsukiyama and Wu, 1995) and later found to promote transcription from chromatin templates *in vitro* (Mizuguchi et al., 1997). BPTF itself is essential for early mouse embryo development, and both BPTF mutant embryos and ESCs show severe defects in gene expression (Landry et al., 2008). Although NURF was already known to be recruited to chromatin by gene-specific transcription factors (Badenhorst et al., 2005), Wysocka et al show that it is stabilized at these loci through its binding to H3K4me3 and that this stabilization is crucial for proper Hox gene regulation and, consequently, development in *Xenopus* embryos. The ING2 protein is a tumor suppressor protein and a native subunit of the mSin3a–HDAC1 histone deacetylase complex, a complex with gene repressive activity (Shi et al., 2006). The discovery that the PHD finger of ING2 binds H3K4me3 and functions to

stabilize a repressive complex on chromatin revealed a novel role of the H3K4me3 mark, which is normally associated with gene activation. It also suggested that H3K4me3 may play a key role in tumor suppressor pathways. Together, the work on the PHD fingers of BPTF and ING2 were critical in furthering our understanding as to how H3K4 methylation is mechanistically linked to gene expression.

An equal important finding was the identification of the chromodomain (CD) as the methyl-binding motif in the chromatin-associated protein Heterochromatin Protein 1 (HP1) (Bannister et al., 2001; Lachner et al., 2001). Like the histone H3K9 HMTase Suvar39 described above, HP1 was originally discovered as Su(var)205 in *Drosophila* (Eissenberg et al., 1990). Work by Eissenberg et al revealed that this gene encodes the heterochromatin-associated protein HP1 and that it enhances PEV in a dosage-dependent manner when expressed exogenously (Eissenberg et al., 1990; Eissenberg et al., 1992). Given that Suvar39, Hp1, and Pc are all enriched at heterochromatin and also contain the evolutionarily conserved chromodomain, Lachner et al set out to test their affinity for H3K9 methylation. Their work and that of others (Bannister et al., 2001) confirmed that the chromodomain region of HP1 binds preferentially to H3K9 methylation, and subsequent work further characterized the structure and affinity of this binding (Jacobs and Khorasanizadeh, 2002; Jacobs et al., 2001). The CD is part of the Royal Superfamily of protein folds, which also includes the Tudor, Chromo barrel, and MBT domains (Taverna et al., 2007). Interestingly, both Royal Superfamily folds and PHD fingers use a common structural feature, a “cage” of aromatic residues, to bind methylated histones. However, subtle differences between them are able to confer sequence and methylation state specificity.

For example, although the CD-containing Pc protein did not show strong affinity for H3K9 methyl in the aforementioned studies, Fischle et al demonstrate that the CD of Pc preferentially binds methylated H3K27 (Fischle et al., 2003b). Moreover, they show that the CDs of *Drosophila* proteins HP1 and Pc discriminate

between histone H3 lysine methylation at K9 (preferentially recognized by HP1) and K27 (preferentially recognized by Pc). Although it had been suggested that the CD of Pc might bind either H3K9me or H3K27me (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Muller et al., 2002), their study was the first to show that the CD of Pc preferentially binds H3K27me over H3K9me. Moreover, they demonstrated the importance of the difference between the CDs of HP1 and Pc in terms of their ability to localize to distinct regions of chromatin *in vivo*. This finding is particularly significant given that PcG proteins are known to play an important role in regulating development and differentiation (see above). PcG proteins are grouped into two Polycomb Repressive Complexes (PRC), PRC1 and PRC2. Several chromatin-immunoprecipitation (ChIP) studies using antibodies against both H3K27me3 and PcG proteins suggest that PRC2 is indeed necessary for the recruitment of PRC1 (Boyer et al., 2006; Bracken et al., 2006; Lee et al., 2006). The H3K27me3 methylase EZH2 is a member of PRC2, and the CD-containing Pc/CBX proteins are members of PRC1 (Whitcomb et al., 2007). The binding of H3K27me3 by the CD of Pc/CBX therefore fits with the model in which H3K27 methylation by PRC2 enables the recruitment of PRC1 (see **Figure 1.2B**).

Another Royal Superfamily member that binds methyated lysines and is found in a group of protein factors that function as transcriptional repressors is the Malignant Brain Tumor (MBT) repeat (Taverna et al., 2007). One of these factors is the PcG protein human Lethal-3 Malignant Brain Tumor repeat-Like protein 1 (L3MBTL1). Consistent with its classification as a PcG protein, L3MBTL1 has been shown to repress transcription in human cells (Boccuni et al., 2003) and its *Drosophila* homologue bears high homology to another *Drosophila* protein encoded by the sex comb on midleg (Scm) PcG gene (Bornemann et al., 1996). Homozygous mutation of the *Drosophila l(3)mbt* gene causes malignant growth of the adult optic neuroblasts and ganglion mother cells in the larval brain, classifying

it as a tumor suppressor gene as well (Gateff et al., 1993; Wismar et al., 1995). A common deleted region of the locus to which the human L(3)MBT gene maps is also associated with myeloid hematopoietic malignancies (Bench et al., 2000; Kurtin et al., 1996; MacGrogan et al., 2001; Wang et al., 2000). Despite this accumulation of functional information about MBT repeat containing proteins, the role of the domain itself was not uncovered until more recently when it was identified as a potential methyl-binding motif (Kim et al., 2006; Trojer et al., 2007). Different from most of the other methyl-binding motifs identified and discussed above, MBT repeats prefer to engage lower methylation states (i.e. mono and di), and are one of the few motifs identified to do so (Taverna et al., 2007). Both the MBT repeats of L3MBTL1 and those of another *Drosophila* PcG protein, Scm-related gene containing Four MBT domains (SFMBT), prefer mono and dimethylated lysines and discriminate against trimethylated lysines (Klymenko et al., 2006; Li et al., 2007).

In Chapter 2, I describe my contribution to the understanding of the molecular basis for methyl-lysine recognition, including the binding of H3K9 and H3K27 methylation by the CDs of the five mammalian Pc proteins, H3K4me3 recognition by the PHD finger motif of BPTF, and lower lysine methyl state recognition by the MBT repeats of the L3MBTL1 protein.

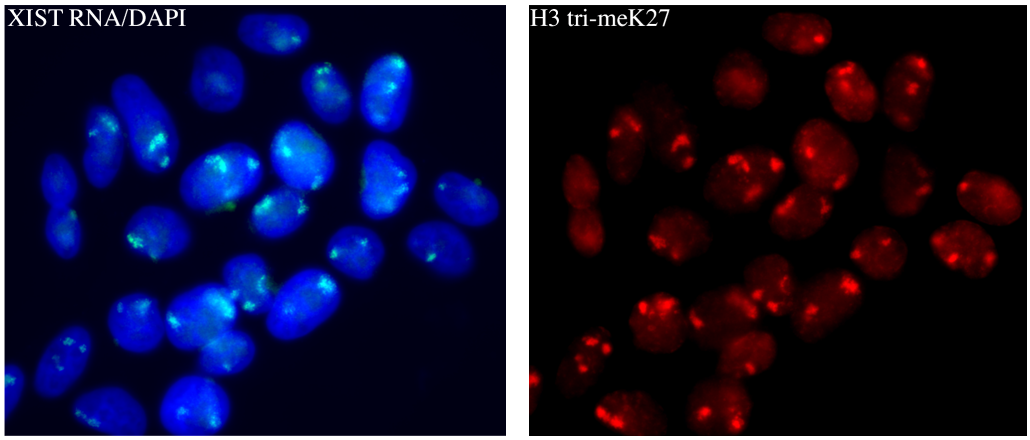
Polycomb Proteins and X Inactivation

In addition to their role in repressing Hox genes, PcG proteins are also implicated in the silencing of the mammalian inactive X chromosome (Heard, 2004). In mammals, female embryonic cells compensate for having two X chromosomes (as opposed to the single X in their male counterparts) by inactivating one of them. The inactive X chromosome (X_i) was identified over 50 years ago as a heterochromatic structure that was only present in female cells (Barr and Bertram, 1949; Ohno et al., 1960). Mary Lyon then proposed that this structure must represent a genetically inactive female chromosome that is silenced early in development and inherited

through mitosis, thus causing the mosaic coat of female mammals such as the calico cat (Lyon, 1961). Subsequent studies have shown that X_i silencing requires a long, non-coding RNA called Xist, which coats the chosen X_i as it transcribes and presumably antagonizes a series of events that lead to stable, heritable inactivation (see Xist RNA Fluorescence In Situ Hybridization (FISH) in **Figure 1.3A, left**). The X_i is categorized as facultative heterochromatin, since its silent status is caused by and dependent upon its environment (namely, the existence of another X in that cell); this dependence on context also implies that the silencing is reversible under certain developmental circumstances. Together the heritability and reversibility of its silent state make the X_i an excellent model for studying the formation of heterochromatin, particularly since it is regulated by cellular events that coincide with and/or result from differentiation (**Figure 1.3B**).

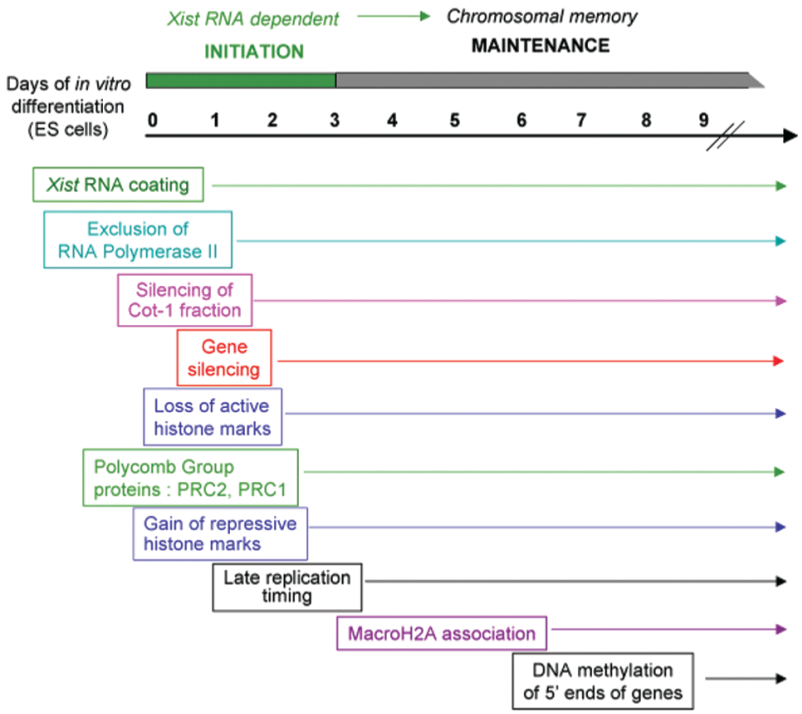
Of interest to our laboratory and many other chromatin laboratories, the silent X chromosome bears a pattern of histone modifications that is characteristic of heterochromatin: hypoacetylation, hypo-H3K4me3, hyper-H3K9me2 and, in particular, hyper-H3K27me3 (see **Figure 1.3A, right**). The mammalian PcG protein EZH2 is a component of PRC2 and is the HMTase responsible for H3K27 methylation on the X_i (Plath et al., 2003); mammalian PcG proteins Ring1a and b are members of PRC1 and has been shown to ubiquinate the X_i on H2A at K119 (de Napoles et al., 2004). However, there are still many open questions about the roles of PcG proteins in X chromosome silencing. For example, although the CD protein dmPc has been extensively studied in the fly, mammals have five different homologues to dmPc, CBX2, 4, 6, 7, and 8, suggesting that they each may behave quite differently. In Chapter 2, I describe collaborative work on the five mammalian CBX proteins and their localization on the X_i during female mouse ESC differentiation. This work introduced me to the mouse ESC model, which I then used for other studies.

A



data produced by Edith Heard and used with permission

B



from Masui O & Heard E. Cold Spring Harb Symp Quant Biol. 2006

Figure 1.3: The mammalian inactive X chromosome is a model of heterochromatin formation and epigenetic silencing.

A. Left panel: RNA Fluorescence *In Situ* Hybridization (FISH) with an XIST RNA probe (green) localizes to the inactive X chromosome in female mammalian cells (HEK-293); the DNA is stained with DAPI (blue). Right panel: immunofluorescent (IF) staining of the same cells with an antibody to the H3K27me3 modification (red); note that the green XIST loci mirror the red H3K27me3 loci, since both are markers of the inactive X chromosome. (Images produced by Edith Heard, Curie Institute, and used with permission.)

B. A timeline of the many events that occur during the process of X-inactivation in differentiating ESCs. (Image from (Heard et al., 2004).)

Chromatin Remodelling During ESC Differentiation

In order to address the question of how the genome of a single cell directs the development of an entire multi-cellular organism, it is important to study the very early stages of the process. The initial stages of mammalian development are difficult to study because the early embryo is composed of very few cells. The development of the Embryonic Stem Cell (ESC) model was a huge step forward in our ability to study those first events in mammalian early embryonic development (Evans and Kaufman, 1981; Martin, 1981; Thomson et al., 1998). After fertilization, the mouse egg undergoes a series of divisions and migrations that results in a hollow cavity of cells, called the blastocyst, after approximately four days (Alberts et al., 2002). As depicted in **Figure 1.4A**, ESC lines are derived from a cluster of cells called the Inner Cell Mass (ICM) that are located on one side of this blastocyst. These cells are termed pluripotent because they can differentiate into all the cell types of the embryo proper. After they are derived, we can test the pluripotency of ESCs by injecting them back into a normal mouse blastocyst and asking whether they contribute to the embryonic tissues. ESC pluripotency is maintained *in vitro* by growing them on either a layer of irradiated (and thus non-dividing) mouse "feeder" cells or plates coated with gelatin. Leukemic Inhibitory Factor (LIF) is also added to the cell media to maintain activation of the JAK-STAT and MAPK pathways, which promotes their proliferation and self-renewal and prevents them from differentiating (Smith, 2001). Removing LIF and/or adding inducers such as retinoic acid leads to ESC differentiation and the loss of pluripotency, causing the cells to undergo dramatic changes in morphology (**Figure 1.4B**), cell cycle, and gene expression (Kim et al., 2008; Murry and Keller, 2008). The ESC model is therefore a tractable system in which to study events that may occur in early embryogenesis when only a few cells are dividing.

Since eukaryotic genomes are intimately associated with histone proteins to form chromatin, this physiologically-relevant structure must be remodeled as part of

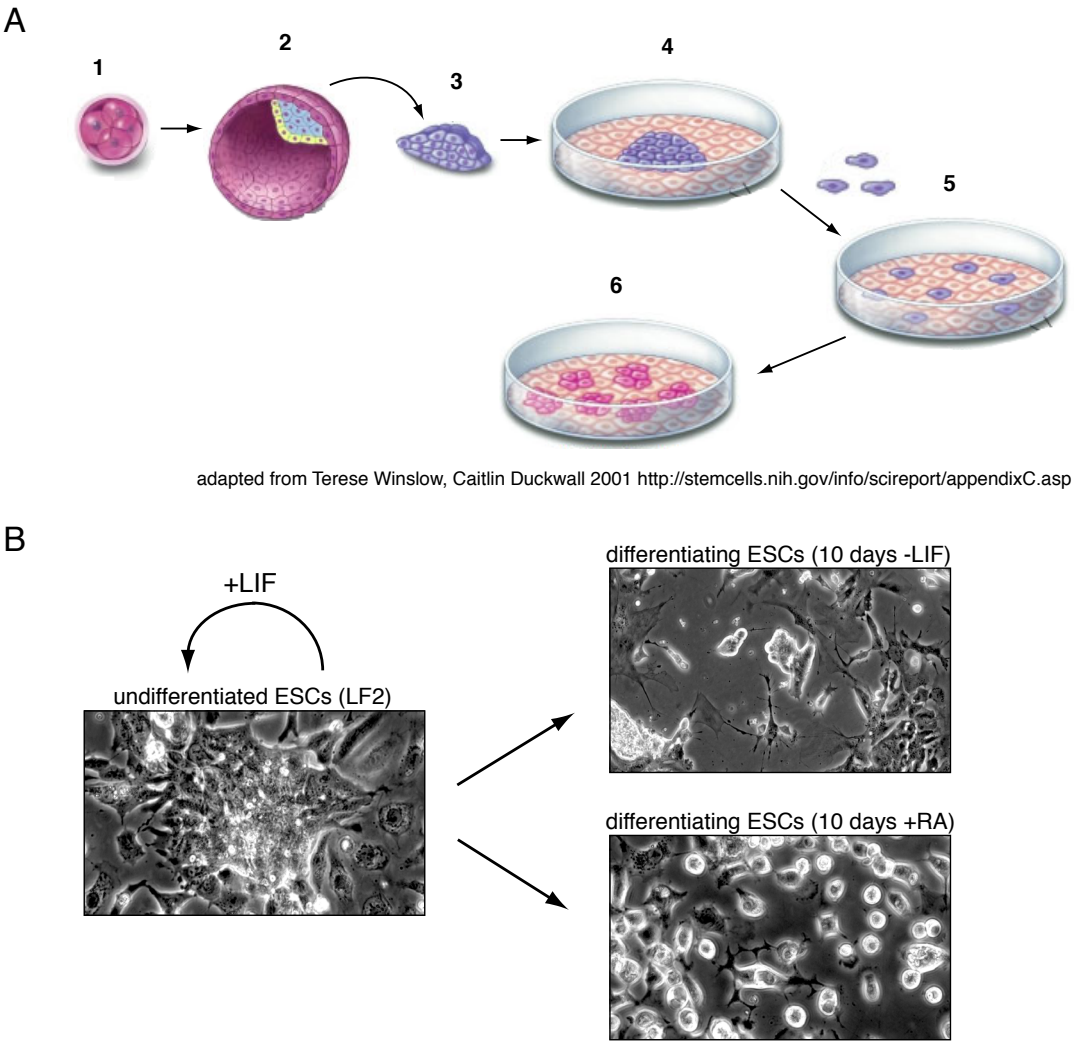


Figure 1.4: Mouse Embryonic Stem Cells (ESCs) are a tractable model of early embryonic development.

A. A schematic of embryonic stem cell (ESC) derivation: the fertilized mouse egg goes through a series of cell divisions (1); after about 4 days, the cells have divided and arranged themselves into a hollow ball called the blastocyst (2); a cluster of cells located on one side of the blastocyst, called the Inner Cell Mass (ICM) is isolated (3) and removed to a plate coated with a layer of mouse feeder cells (4); after a period of recovery and growth, cells are split and re-plated onto new feeder cells (5) or plates coated with gelatin and allowed to grow up into new colonies(6). (Image adapted from (Winslow and Duckwall, 2001).)

B. Phase-contrast images of live ESCs taken with the Olympus IX71 inverted microscope in the Bio-Imaging Center at Rockefeller University. The self-renewal property of ESCs is maintained in the presence of Leukemic Inhibitory Factor (LIF); undifferentiated ESCs are highly proliferative and prefer to grow in tight, dense clusters (left). After 10 days of differentiation with either LIF removal alone (right, top) or LIF removal + Retinoic Acid (RA, right, bottom) the cells have changed in morphology, becoming longer and more spindle-like in some cases and rounder and flatter in others; some cells in under differentiation conditions appear to make fewer cell-to-cell contacts and extend neuronal-like processes.

a large-scale mechanism to achieve rapid and drastic changes in gene expression (Arney and Fisher, 2004; Gan et al., 2007). For example, undifferentiated ESCs typically display increased physical plasticity and less compacted chromatin than their differentiated counterparts (Meshorer et al., 2006; Pajerowski et al., 2007). ESCs also undergo radical changes in gene expression as they differentiate, conveniently providing markers of "stemness" whose expression dramatically decreases (e.g. Oct 3/4) as differentiation progresses. Such changes suggest that cells undergo a significant reorganization of their genome during the differentiation process and that, moreover, this transition must be carefully regulated in order for the cell to differentiate properly and adopt a specific lineage.

Recent studies have shown that histone covalent modification patterns change significantly upon ESC differentiation (Giadrossi et al., 2007). For example, core histones (H2A, H2B, H3 and H4) are largely deacetylated upon differentiation and histone deacetylase activity may be required for ESC differentiation (Lee et al., 2004). Chromatin-immunoprecipitation (ChIP) experiments have also identified specific genes and/or genomic regions that change their "epigenetic signature" upon differentiation (Azuara et al., 2006; Bernstein et al., 2006). Specifically, these researchers found that certain genes in undifferentiated ESCs were uniquely modified with both H3K4me3, a marker of gene activation, and H3K27me3, a marker of gene repression. This finding was novel in that it not only suggested that ESCs have a unique chromatin biochemical profile, but also that the paradigm of active genes being marked by H3K4me3 and silent genes by H3K27me3 does not always hold. However, this "bivalent domain" signature, as they termed it, resolves to the usual paradigm upon differentiation. In Chapter 3, I describe my own Chromatin ImmunoPrecipitation (ChIP) studies in ESCs using an antibody raised against the H3K4me3 modification and show that this model fits with other descriptions of changes in chromatin during ESC differentiation.

Remodelling Chromatin by Histone Isoform Replacement

Despite a wealth of emerging data describing changing patterns of epigenetic signatures during ESC differentiation, very little is known about the mechanisms used to achieve such change. Several possible mechanisms for removing the more stable histone modifications, e.g. lysine methylation, include enzymatic demethylation, histone replacement, and regulated histone proteolysis (**see Figure 1.5A**) (Bannister and Kouzarides, 2004). Although the identification of lysine demethylases was enigmatic for many years, the discovery that LSD1 removes mono and dimethyl groups from histone H3K4 (Shi et al., 2004) began a cascade of findings about this type of enzyme. Since then, many demethylating enzymes have been identified, including those that can remove trimethyl groups and at least one that can remove methylation from arginine residues (Lan et al., 2008).

Similarly, for many years there was evidence that histone replacement occurred outside of S-phase, yet the mechanism was not know. Although the bulk of histone deposition into nucleosomes occurs during S-phase as DNA is replicated, studies in rat neurons (Pina and Suau, 1987), mammalian tissues (Lennox and Cohen, 1988) and *Tetrahymena* (Yu and Gorovsky, 1997) show that variant isoforms of core histone proteins are incorporated into the chromatin fiber independent of replication. Histone variant incorporation, specifically H3.3, was then first linked to actively transcribing loci in *Drosophila* (Ahmad and Henikoff, 2002). Ahmad and Henikoff demonstrate that the alternate histone H3 variant, H3.3, is specifically and exclusively deposited at active loci outside of S-phase despite the fact that it differs in only four amino acids from the canonical *Drosophila* histone H3 (**Figure 1.5B**). Although typical replication-dependent histone deposition requires the N-terminal tail of H3, Ahmad and Henikoff further show that replication-independent deposition does not require the N-terminal tail. Subsequent studies by Tagami et al then identified the enzymatic machinery responsible for these distinct deposition pathways when they purified two separate chaperone complexes with H3.1 (i.e.

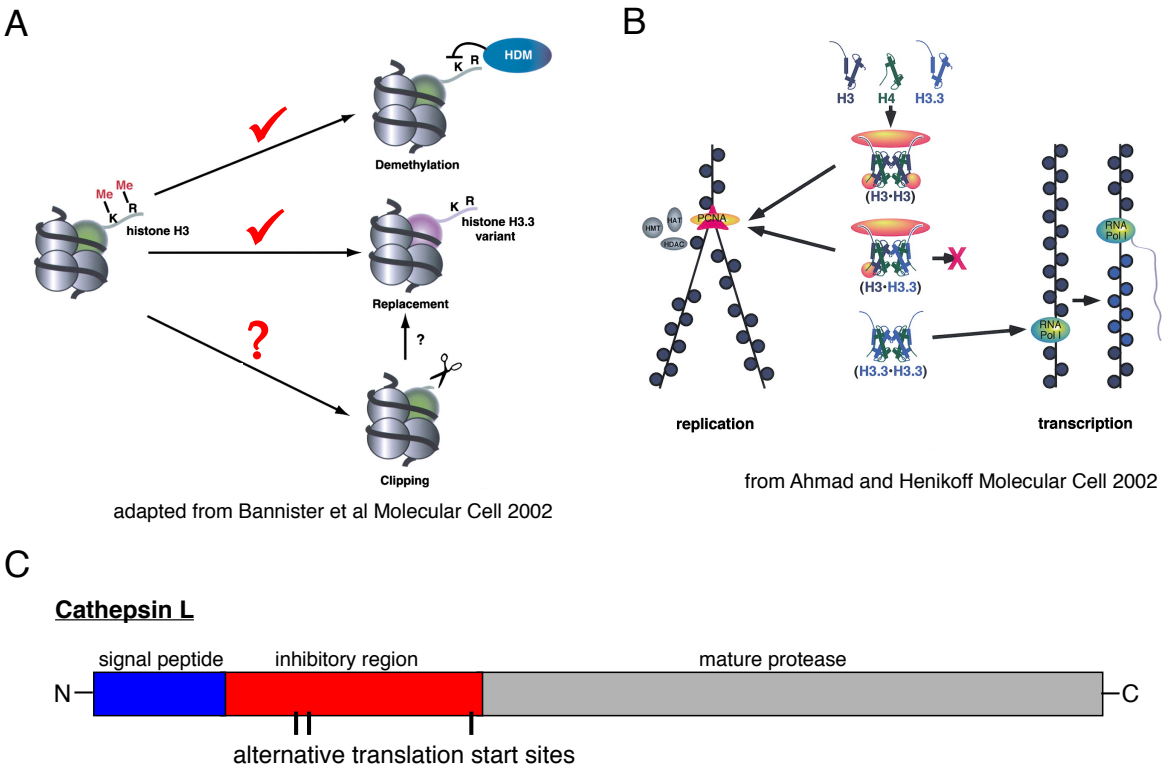


Figure 1.5: Reversible and irreversible mechanisms for removing histone tail modifications.

A. Suggested mechanisms for the removal of methyl marks from chromatin: enzymatic demethylation, regulated histone replacement (including outside of S-phase), and histone tail "clipping" or proteolysis. The red "check" marks indicate that the first two mechanisms have been well documented in the literature. The red question mark indicates that histone clipping has not yet been established as a mechanism to regulate the removal of histone modifications. The vertical arrow and black question mark suggests that histone clipping may lead to replacement of the entire histone protein. (Image adapted from (Bannister and Kouzarides, 2004).)

B. A schematic of replication-dependent (left) versus replication-independent (right) histone deposition. Ahmad and Henikoff show that the latter couples the incorporation of the histone H3 variant isoform H3.3 with active transcription (right). The red "X" indicates that H3.3 (variant) + H3 (canonical) heterodimers are not deposited into chromatin. (Image taken from (Ahmad and Henikoff, 2002).)

C. Cathepsin L, like most cathepsins, is synthesized as a preproprotein: the signal peptide (blue) is removed in the endoplasmic reticulum, and the inhibitory region (red) must be removed in order for the mature protein (grey) to be activated. The Cathepsin L protein sequence contains alternative translation start sites that would exclude the signal peptide and, for at least one site, most of the inhibitory region.

canonical H3) and H3.3 from mammalian cells (Tagami et al., 2004). In Chapter 3, I describe collaborative work in which we provide evidence that suggests regulated histone H3 replacement occurs during ESC differentiation.

Remodelling Chromatin by Regulated Proteolysis

There is also precedence for the third aforementioned mechanism of histone modification removal, controlled histone proteolysis, in both *Tetrahymena thermophila* and mammalian cells infected with the foot-and-mouth disease virus (Allis et al., 1980; Falk et al., 1990). Both these studies describe H3-specific histone proteolysis. Recent work in the Kouzarides laboratory also suggests that histone H3 cleavage occurs in *s.cerevisiae* (Santos-Rosa et al., 2009). However, specific, regulated, endogenous proteolysis of histone proteins had not been well documented in mammalian cells. In Chapter 4, I describe my characterization of H3 proteolysis in differentiating ESCs and identification of Cathepsin L as a protease that creates this cleavage.

Cathepsins are a subclass of cysteine proteases in the papain family; the majority have only endopeptidase activity, although there are some exceptions (Barrett et al., 2003). Cathepsins are normally synthesized as preproenzymes (**Figure 1.5C**), including a signal peptide that directs them to the endoplasmic reticulum and an inhibitory region that must be cleaved for their activation (Barrett et al., 2003). The removal of this inhibitory region is thought to occur via autocatalytic processing at acidic pH or by the peptidase activity of other proteases, such as the aspartic protease Cathepsin D (Barrett et al., 2003). Cathepsins L and B are the most abundant lysosomal cysteine proteases, and the double knockout of both genes is lethal during the second to forth week in the mouse (Felbor et al., 2002). Interestingly, the single knockout of either gene is not lethal but rather produces divergent phenotypes, suggesting that these two proteases have both overlapping and unique roles (Turk et al., 2000).

Although cathepsins were previously thought to reside exclusively in the lysosome, recent evidence suggests that cathepsins, in particular Cathepsin L, may localize to the nucleus under certain conditions (Goulet et al., 2004; Hiwasa and Sakiyama, 1996). Despite the obvious differences in environment between the lysosome and the nucleus, Goulet et al show that Cathepsin L has limited proteolytic activity when localized to the nucleus and plays a role in regulating transcription factor binding and cell cycle progression (Goulet et al., 2004). Their data suggesting that the nuclear form of Cathepsin L is translated from an alternative start site downstream of the signal peptide. Moreover, Goulet et al propose that the non-optimal nuclear pH may be an important feature in the regulation of Cathepsin L activity in the nucleus. Other biochemical studies have also shown that pro-Cathelsin L is localized to the nucleus in *ras*-transformed mouse fibroblasts (Hiwasa and Sakiyama, 1996). In addition, evidence of a naturally occurring nuclear serpin inhibitor (Myeloid and Erythroid Nuclear Termination stage-specific protein, MENT) with the ability to inhibit Cathepsin L activity (Bulynko et al., 2006; Irving et al., 2002) further supports the notion that Cathepsin L and other cysteine proteases do indeed traffic to the nucleus to play important roles in regulated nuclear proteolysis.

Such regulated proteolysis of histone proteins could have broad impact on chromatin structure and function. For example, removing the histone H3 tail also removes the post-translational modifications on that sequence and any downstream effectors recruited by those marks. Furthermore, as I will describe in Chapters 2 and 4, the removal of the N-terminal tail also impacts the binding affinity of proteins, such as Pc, that bind modifications C-terminal to the cleavage site. Taken together, it is clear that the changes that occur within chromatin during differentiation and development are both dynamic and mutually dependent.

Chapter 2: Biophysical studies of methyl-histone binding proteins

Summary

One major goal of our laboratory is to understand the downstream biological effects of covalent histone modifications. In this chapter, I describe three distinct, yet mechanistically related, projects in which I investigated the biophysical interactions between methylated histone proteins and several non-histone proteins. In collaboration with the Heard Laboratory at the Curie Institute, post-doctoral fellow Emily Bernstein and I investigated the binding interactions of all five mammalian Polycomb proteins with methylated histones and the role of this interaction *in vivo* (Bernstein et al., 2006). After a colleague in the Allis laboratory discovered that a Plant HomeoDomain (PHD)-finger of the NUcleosome Remodeling Factor (NURF) chromatin-remodeling complex binds to histone H3K4me3 (Wysocka et al., 2006), I collaborated with the Patel Laboratory at Memorial Sloan Kettering Cancer Center to investigate the molecular basis of this interaction (Li et al., 2006). In another collaboration with the Patel Laboratory, I investigated the differences in molecular interaction between the trimethyl-lysine binding PHD finger and the mono- and dimethyl-lysine binding Malignant Brain Tumor (MBT) repeats of the protein L3MBTL1 (Li et al., 2007).

Results

Mouse Pc Proteins Bind Differentially to Methylated Lysine H3 *In Vitro*

I first started studying the mammalian Polycomb (Pc) proteins as part of a collaborative project with Emily Bernstein, a post-doctoral fellow in our lab. In particular, we focused on the five mammalian homologues of the namesake *Drosophila* Pc protein that contains the defining chromodomain (CD). The chromodomain (CHRomatin Organization MOdifier domain) was originally identified as a conserved sequence motif in proteins that are modifiers of variegation in *Drosophila*, including Pc and Heterochromatin Protein 1 (HP1), both of which induce silencing by condensing chromatin structure (see Chapter 1 and **Figure 1.2A**). The mammalian homologues to both these proteins are named CBX after this domain, otherwise known as a ChromoBoX (HP1=CBX1, 3, 5, Pc=CBX2, 4, 6, 7, 8).

The CDs of both dmHP1 and dmPc have been shown to bind methylated histone H3 (Bannister et al., 2001; Fischle et al., 2003; Jacobs and Khorasanizadeh, 2002; Jacobs et al., 2001; Lachner et al., 2001; Min et al., 2003). Specifically, the CD of *Drosophila* Pc has been shown to have a higher affinity for H3 peptides trimethylated at lysine 27 while dmHP1 strongly prefers those trimethylated at K9 (Fischle et al., 2003). Although the mechanism of chromatin compaction through the recruitment of repressive binding proteins such as HP1 and Pc is still unclear, this differential binding suggests a paradigm in which HP1 is recruited to genomic loci enriched in H3K9me3 and Pc to those loci decorated by K27me3 (**Figure 2.1A**). Since the CDs of the mouse Pc proteins (CBX2, 4, 6, 7, 8) share a high degree of sequence conservation with that of dmPc itself (**Figure 2.1B**), we posited that they would show strong affinity for the H3K27me3 mark as well.

After we cloned and optimized the expression of the chromodomains of these five proteins, I used two *in vitro* binding assays to assess the binding ability of

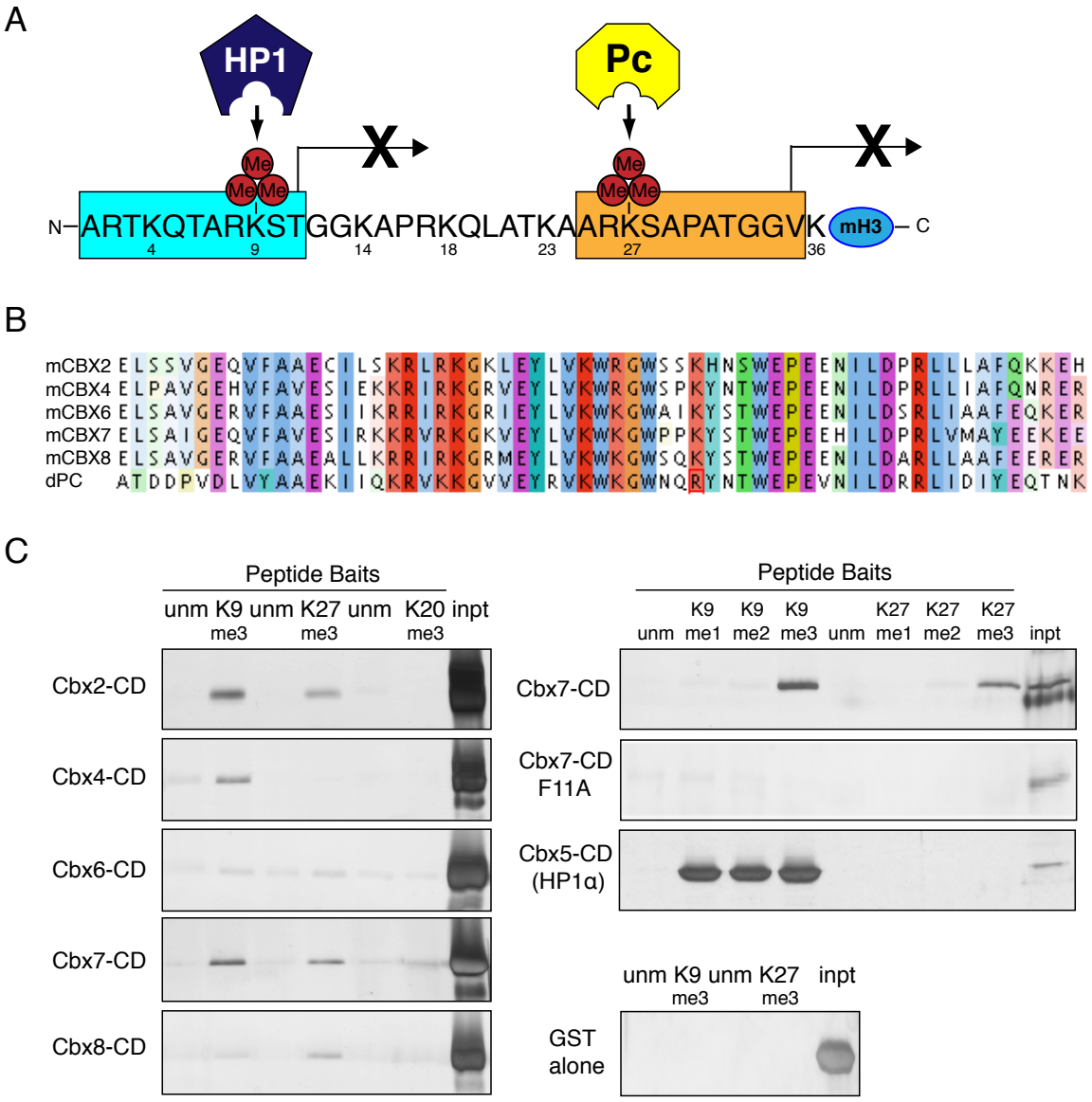


Figure 2.1: Mammalian Polycomb proteins expand the *Drosophila* HP1/Pc binding paradigm.

A. An illustration of the dmHP1/Pc binding paradigm: dmHP1 preferably binds to H3K9me3 while dmPc has a higher affinity for H3K27me3. Their interactions with these histone methyl marks are thought to stabilize these repressor proteins at specific genomic loci as they mediate chromatin compaction and gene silencing, although the exact mechanisms are still unclear.

B. A ClustalW alignment of the five mouse Pc-like chromodomains (CDs; aa 1-62) and the *Drosophila* Pc CD (aa 16-78). Asterisks represent the caging aromatic residues that mediate the histone methyl-lysine interaction. Note the high degree of conservation among these family members.

C. Peptide pull-down assays. Left: CBX2, 4, 6, 7 and 8 were examined for binding to unmodified and trimethylated peptides representing H3K9, H3K27 and H4K20. Right: CBX7 CD, CBX7 caging aromatic point mutant F11A, and human HP1β recombinant proteins were tested for the ability to bind unmodified and me1, 2 and 3 peptides of H3K9 and K27. As a control, GST alone was also tested for binding to unmodified and trimethylated peptides representing H3K9 and H3K27.

these Pc proteins and/or their CDs alone to modified histone tails (e.g. H3 K27m3). First, I performed peptide "pull down" assays in which histone peptides are labeled with biotin, conjugated to avidin-coated beads through the strong interaction of biotin with avidin, and incubated either with cell extracts or, as in the experiment shown in **Figure 2.1C**, a mixture of recombinant protein plus 20 fold excess of BSA. The beads are then washed rigorously and the protein eluted from the beads. The eluates are then run on SDS-PAGE gels and stained. As shown in **Figure 2.1C**, left, CBX2-CD and CBX7-CD both bind strongly to H3K9me3 as well as to H3K27me3; CBX4-CD prefers H3K9me3; CBX8-CD shows a slight preference for H3K27me3 over H3K9me3 or unmodified peptide, and CBX6-CD does not show strong affinity for any of these peptides. I also tested the full-length CBX7 protein in this assay, and it behaves identically to the CD alone (data not shown). We also tested a mutant CBX7-CD protein (F11A), in which one of the three residues that form the aromatic cage of the CD was changed to an alanine (**Figure 2.1C, right**). Earlier mutagenesis and structural studies have shown that the conserved hydrophobic, aromatic pocket, or cage, of both the HP1 and Pc chromodomains is responsible for the recognition of the methylammonium group of methylated lysine K9 and K27, respectively (Jacobs and Khorasanizadeh, 2002; Jacobs et al., 2001; Min et al., 2003). In keeping with what was previously reported, mutating this residue to alanine abolished the binding affinity of CBX7-CD for both H3K9me3 and H3K27me3. I also purified and tested mouse CBX5-CD protein (mHP1α) as a control for preferential H3K9me3 binding, and my results support those using human HP1α (Fischle et al., 2005). Since these proteins all contain GST tags on their N termini, I also purified recombinant GST protein and tested it for binding; as shown in **Figure 2.1C** (right), GST tag alone did not show any affinity for the key peptides used in our assay.

I then quantified the binding affinities of these chromodomains using fluorescence anisotropy (Park and Raines, 2004). In this assay, recombinant protein was incubated with short histone peptides, either unmodified (control) or modified (experimental), fluorescently labeled with a Ferrocenyl substituted AziridinylMethanol (FAM) conjugate. Upon binding with the protein, the movement of the fluorescent peptide is slowed, and the degree of this retardation (its fluorescence anisotropy) is calculated from the polarization measurement of each binding reaction. These numbers are then plotted against the concentration of protein in each reaction well to determine a binding constant (K_d). **Figure 2.2A** shows representative binding curves of CBX7-CD and CBX4-CD with H3 K9me and H3 K27me peptide (mono-, di- and tri-methylated peptides and an unmodified H3 peptide were tested for each) as well as full length CBX7 protein, which produced nearly identical binding curves to those of CBX7-CD. I also tested CBX5-CD, the mouse homologue of HP1α, as a control and my results supported those shown above in the peptide pulldown assay: CBX5-CD bound much more strongly to methylation at H3K9 than at H3K27 (**Figure 2.2B**). The binding constants of each CBX-CD to these peptides are then summarized in **Figure 2.2C**, for comparison. These results support and quantify those of the peptide pulldown assay above: the control, HP1α, shows a 10-fold preference for H3K9me3 over H3K27me3 (2μM vs 204μM), CBX7 binds strongly to H3 K9me3 (12μM) and H3 K27me3 (22μM), and CBX4-CD prefers H3 K9me3 over K27me3 much like HP1 (49μM vs 150μM). Moreover, CBX7-CD binds either H3K9me3 or H3K27me3 with greater affinity than any other Pc-CD, although still not as strongly as HP1.

Mouse Pc Proteins Bind Differentially to the H3K27me3 Enriched X_i *In Vivo*

Together, the data shown in **Figure 2.2** challenges the HP1/Pc paradigm established with the *Drosophila* proteins (**Figure 2.1A**): although the CD of dmPc strongly prefers H3 K27me3, the CDs of both mouse CBX2 and CBX7 have affinity for both

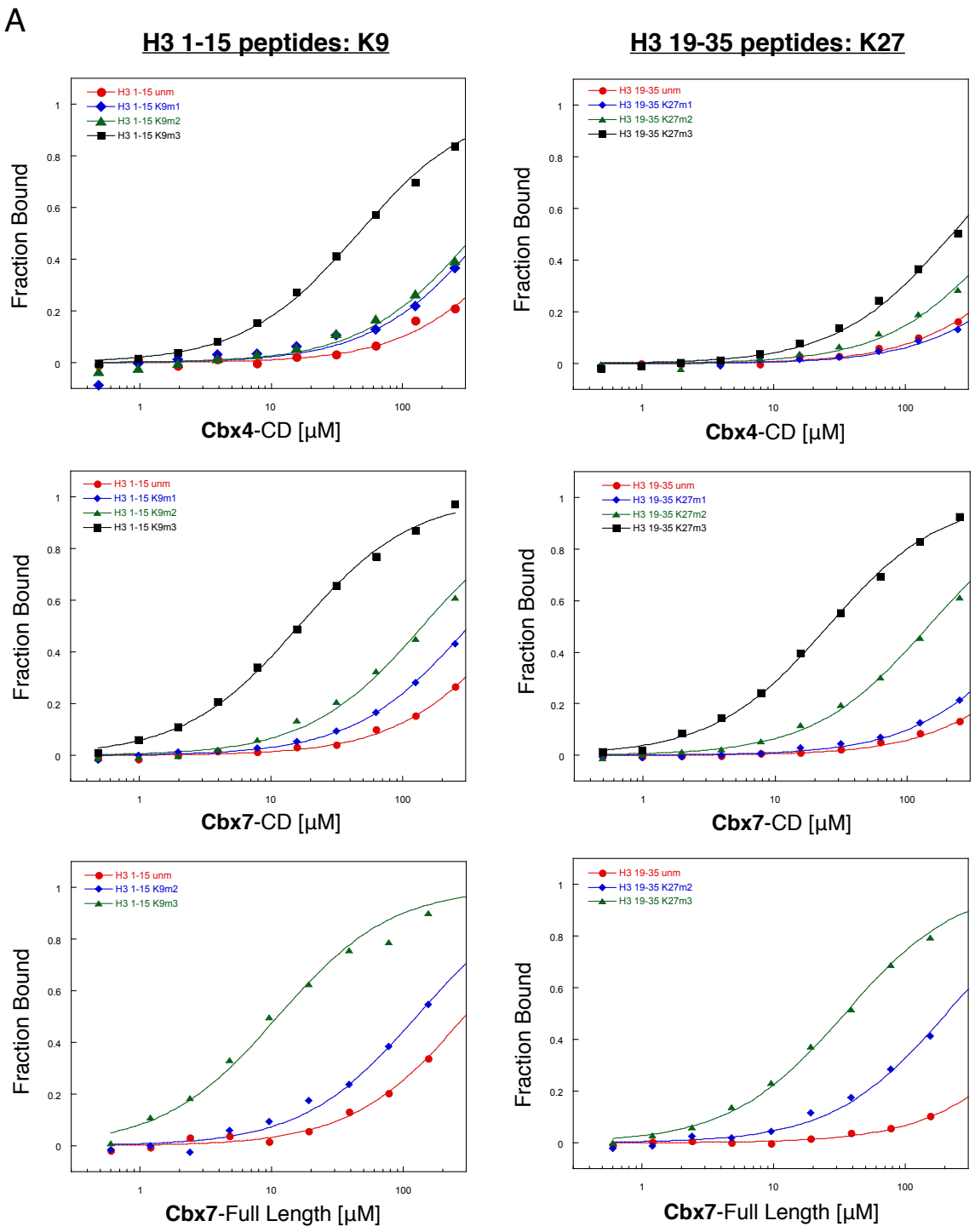
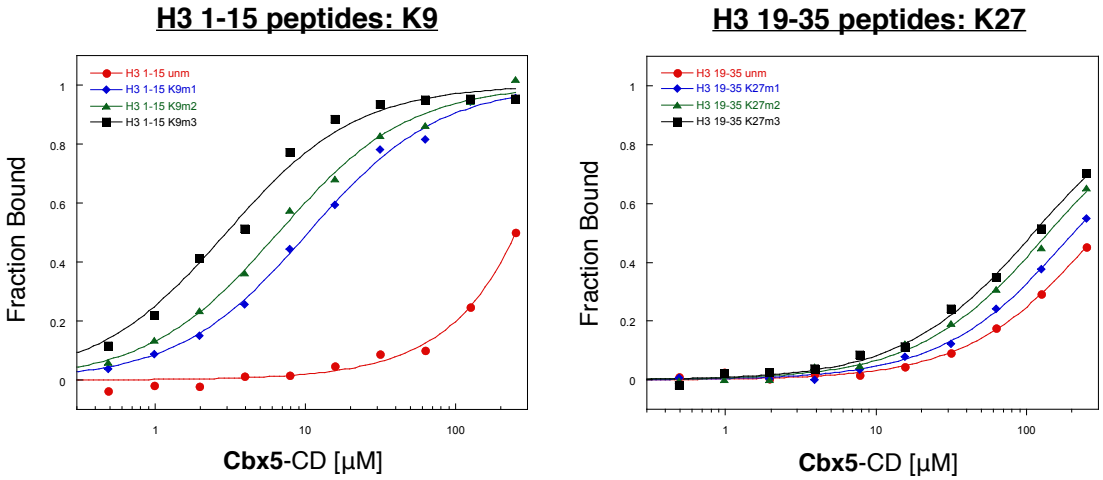


Figure 2.2: Quantification of the mammalian Pc-CDs binding affinity for methylated H3 peptides.

B



C

	H3 1-15 peptides				H3 19-35 peptides		
	unmod	**K9m1**	**K9m2**	**K9m3**	**K27m1**	**K27m2**	**K27m3**
CBX2-CD	>500	382±29	396±36	41±6	>500	143±4	44±5
CBX4-CD	>500	>500	261±21	49±9	>500	>500	150±20
CBX6-CD	>500	>500	>500	>500	>500	>500	330±120
CBX7-CD	>500	267±48	79±12	12±3	>500	136±23	22±5
CBX8-CD	>500	>500	>500	>500	>500	>500	165±20
CBX5-CD (HP1α)	>500	10	7	2±1	>500	266	204

Figure 2.2: Quantification of the mammalian Pc-CDs binding affinity for methylated H3 peptides.

A. Representative fluorescence polarization binding curves of CBX4-CD (top), CBX7-CD (middle), and CBX7-Full-Length to histone tail peptides, including the me1, me2 and me3 states on residues K9 (left) and K27 of H3 (right).

B. Representative fluorescence polarization binding curves of Cbx5-CD (mouse HP1α) to H3K9 (left) and H3K27 (right) histone tail peptides in the me1, me2 and me3 states.

C. Dissociation constants (K_D in μM) for Cbx 2, 4, 6, 7 and 8, as well as Cbx5 (mHP1α), with both H3K9 (aa 1-15) and H3K27 (aa 19-35) peptides. Values represent mean ± SD for at least 3 independent experiments in all cases (except for Cbx5 with certain peptides); ND=not determined.

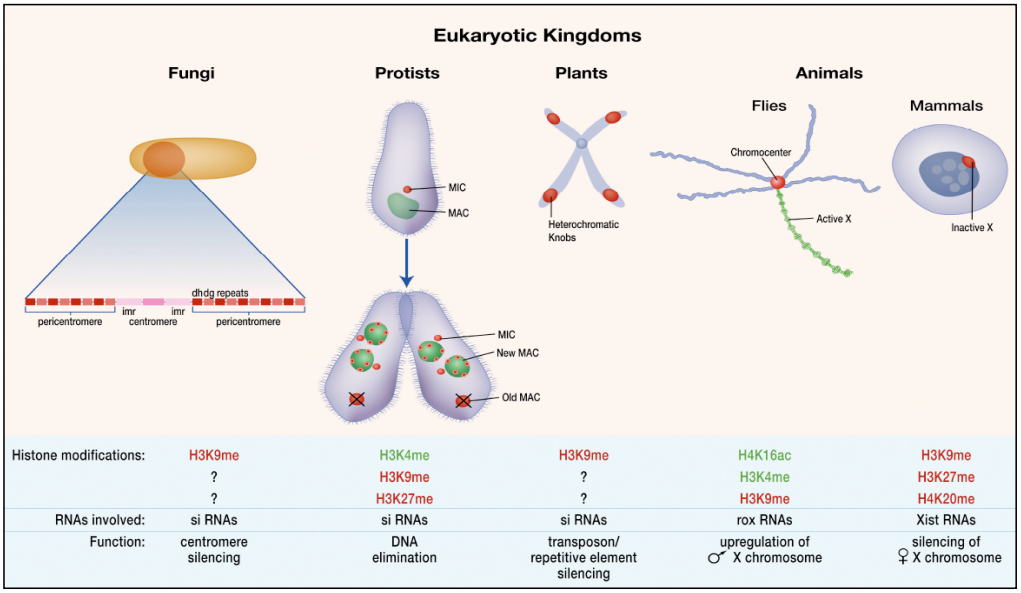
H3 K27me3 and H3 K9me3, and CBX4-CD prefers H3 K9me3. The differential binding preferences of these CDs suggests that 1) the full-length CBX proteins differ from one another in their histone binding affinities and 2) such differences in binding may translate to distinct biological readouts. This finding led us to seek an appropriate *in vivo* model in which to study these Pc-like proteins. Many organisms

use these marks to regulate and/or "remember" gene silencing (see **Figure 2.3A** for depiction), and we chose to study these proteins in the context of their potential role in mammalian X-inactivation.

Mammalian X-inactivation is a classic example of heterochromatin silencing and epigenetic memory. Organisms use different mechanisms to ensure equal dosage of the genes on their sex chromosomes: male *Drosophila* upregulate the transcription of their X-chromosome, *C. elegans* hermaphrodites downregulate each of their two X-chromosomes by half, and female mammals silence one of their two X-chromosomes by converting it to heterochromatin. It is known that the initiation of X-inactivation and establishment of the inactive X chromosome (X_i) requires the non-coding RNA Xist (Penny et al., 1996). However, Xist is only required at a crucial, early window of time; in experiments where Xist was removed after this point, there was no effect on the silent state of the X_i (Csankovszki et al., 1999). Therefore, many scientists hypothesize that the maintenance and clonal inheritance of the X_i must require the recruitment of specific proteins, perhaps including Pc proteins, whose function is to convert the selected X chromosome into its stable heterochromatic state. It is known that trimethylated K27 is a hallmark of the silenced X chromosome in female mammalian cells, and researchers in both the chromatin and X-inactivation fields questioned whether Pc proteins bind this mark as well. If so, the question remains as to how such binding plays a role in the silencing of the chosen X chromosome.

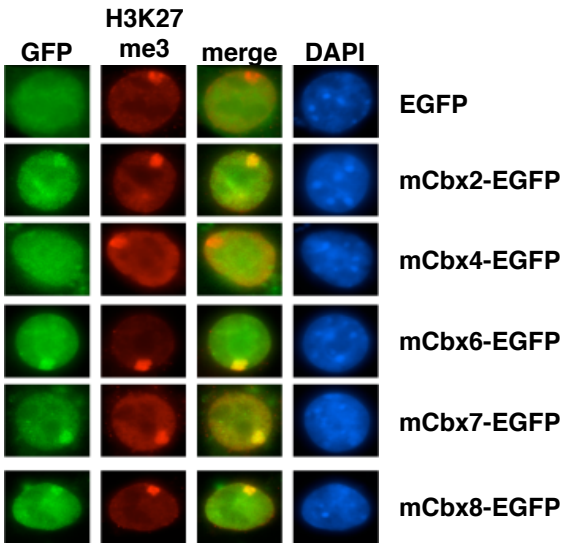
In order to study X inactivation, many labs use a female mouse embryonic stem cell (ESC) differentiation model. We collaborated with the Heard Laboratory at the Curie Institute to study the binding of mouse Pc proteins to the H3K27me3-enriched inactive X in this *in vivo* model. The Heard Laboratory observes the formation of X_is in 3 day retinoic acid (RA) differentiated mouse ESCs and observes an enrichment of H3K27m3 on those X_is by immunofluorescence (IF) combined

A



from Bernstein and Allis, Genes & Dev, 2005

B



*data in B produced by Edith Heard and used with permission

Figure 2.3: *In vivo* association of mammalian Pc proteins with the inactive X chromosome.
A. Model organisms from all eukaryotic kingdoms use mechanisms involving non-coding RNAs and histone modifications to regulate gene expression. In the animal kingdom, both flies and mammals use such mechanisms to regulate gene dosage compensation on the sex chromosomes. In flies, the roX RNAs are integral to the two-fold upregulation of transcription on the male X chromosome and it is marked by "active" histone modifications, such as H3K4me and H4K16ac. In female mammals, *Xist* RNA is required to silence one of two X chromosomes, which is also marked by "silencing" marks (H3K9me, H3K27me, and H4K20me). "Active" chromatin is represented in green shading and "silent" chromatin is shown in red. (Image taken from (Bernstein and Allis, 2005).)
B. Immunofluoresce (IF) images of CBX-EGFP fusion proteins in 3 day +RA differentiated female ESCs (green). All CBX-EGFP proteins, except CBX4, localize to the Xi. The Xi is visualized by K27me3 staining (red) and DNA is stained with DAPI (blue). (Images produced by Edith Heard and used with permission.)

with RNA-FISH for Xist (as originally described in (Plath et al., 2003). In order to test whether or not the five mammalian Pc proteins also co-localize to the X_i, Heard and colleagues transfected GFP-fusion constructs into differentiating ES cells and then used IF to stain the X_i for H3 K27me3 (unfortunately, the available α-CBX abs do not work in IF). As shown in **Figure 2.3B**, several of these CBX-GFP fusion proteins co-localize with K27me3 staining, suggesting that they are recruited to the X_i. Albeit correlatively, this data supports the X_i binding data in **Figure 2.2**: CBX7 is both the strongest H3 K27m3 peptide binder *in vitro* and that which co-localizes with H3 K27me3 on the X_i most robustly.

CBX7 Associates with Chromatin Upon Differentiation in ESCs

Nevertheless, while the *in vitro* binding and GFP-fusion protein localization experiments described above provide information as to whether or not specific associations are possible, they do not directly demonstrate if or when these associations actually occur *in vivo* during ESC differentiation. Although our initial interest in the CBX/Pc proteins grew from a collaborative project that aimed to analyze their potential role in X-inactivation, the lack of IF-compatible antibodies to these proteins makes it impossible to assess the kinetics of their endogenous association with the X_i. For this reason, we decided to survey the association of CBX proteins with chromatin at a global level. To address this question, I isolated chromatin from ES cells at various time points of differentiation (see **Figure 2.4A** for schematic of procedure) and surveyed for changes in the presence of CBX proteins (using antibodies that have been shown to work for immunoblotting). As shown in **Figure 2.4B**, CBX7 showed a dramatic difference in chromatin association between undifferentiated and differentiated ES cells. Although the protein is present throughout the time course, it is clearly only associated with the chromatin (P3) of the differentiated cell samples, starting from day 6 (in **Figure 2.4B,** compare CBX7

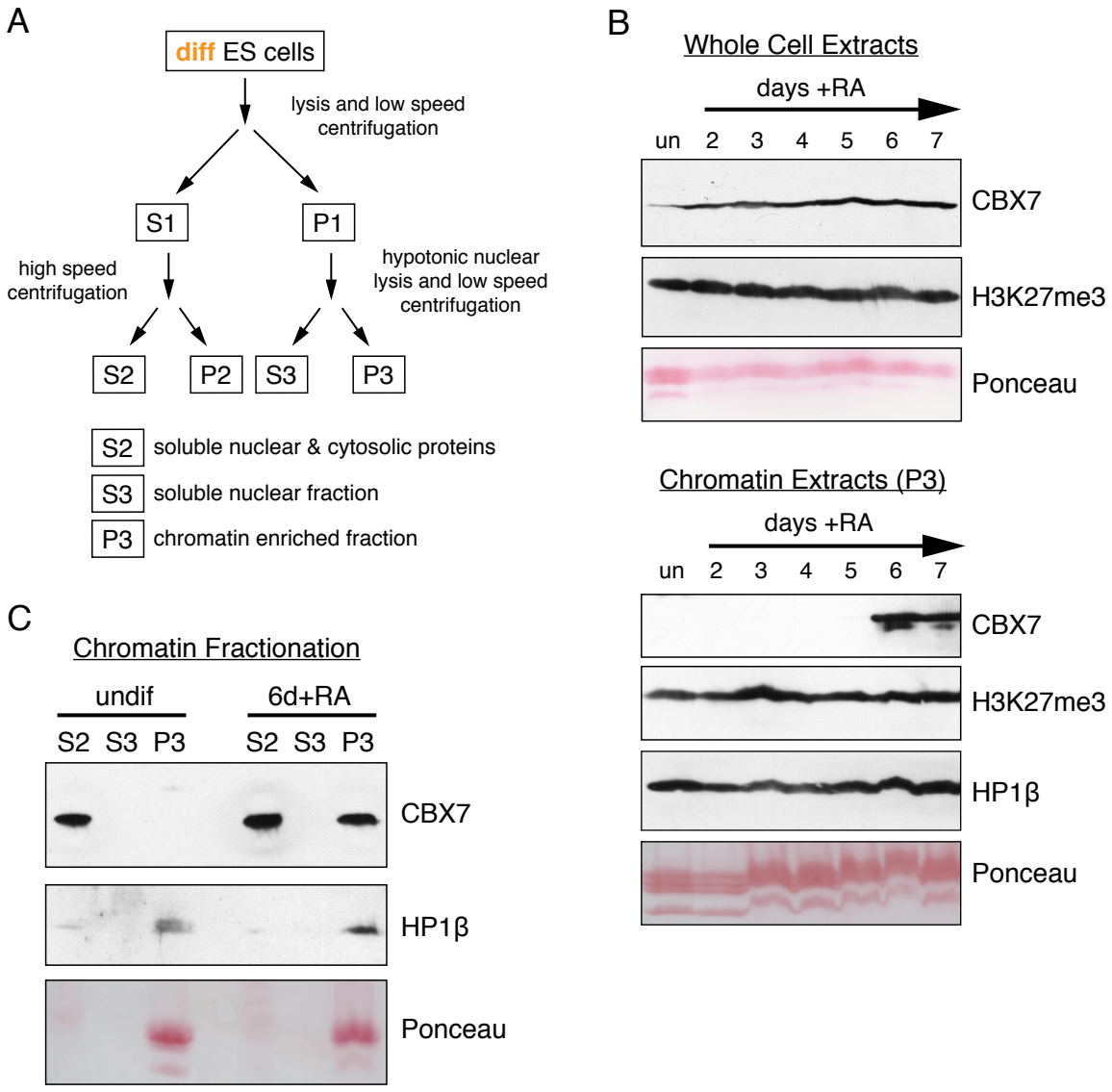


Figure 2.4: Differential global chromatin-association of mammalian Pc protein CBX7.

A. A schematic of chromatin fractionation by centrifugation.

B. Female ESCs were tested for CBX7 chromatin-association through day 7 of differentiation (un=undifferentiated). CBX7 and K27me3 are expressed throughout the differentiation process (top, "Whole Cell Extracts"), yet CBX7 only associates with chromatin ("Chromatin Extracts (P3)") on days 6 and 7 and is absent from chromatin prior to day 6. In contrast, both HP1β and K27me3 are associated with chromatin throughout the differentiation time course. The PVDF membranes were stained with Ponceau to demonstrate that samples were loaded in roughly equal amounts.

C. Undifferentiated (un) and 6 days +RA differentiating ESCs were fractionated according to the scheme in **A**. CBX7 protein is present in the S2 fraction of both undifferentiated and 6 days differentiating cells, but is only present in the P3 fraction of the 6 days differentiateing cells. Another repressive chromatin protein, HP1, does not change its global chromatin associated in the same samples. The PVDF membrane was stained with Ponceau to demonstrate that samples were loaded in roughly equal amounts.

panel of “Whole-cell extracts” versus that of “Chromatin-extracts”). On the other hand, when the same blot was stripped and probed with an antibody against HP1β, which also binds silent chromatin marks, there was no difference in chromatin association (**Figure 2.4B**, “Chromatin Extracts”), suggesting that this effect is not general for heterochromatin binding proteins. Moreover, H3K27me3 is maintained through the time course (**Figure 2.4B**, both “Whole Cell Extracts” and “Chromatin Extracts” panels) suggesting that this binding is mediated through additional mechanisms that remain unclear. In **Figure 2.4C**, all chromatin isolation fractions (see schematic **Figure 2.4A**) from undifferentiated and 6 days RA differentiated cells are compared, and data supports that in **Figure 2.4B**.

CBX Proteins Bind RNA *In Vitro* and Show RNA-dependent Chromatin Association *In Vivo*

We then wanted to test whether the CDs of CBX proteins might interact with RNA in addition to H3K27 methylation in order to be recruited and/or stabilized in their association with chromatin. My colleague Emily Bernstein first performed a series of *in vitro* RNA gel-shift assays (REMSAs) to assess whether the CDs of CBX proteins could bind to RNA molecules. Her findings suggested that all CBX-CDs except CBX2-CD bound ssRNA (**Figure 2.5A**), and that the CD of a mammalian Pc protein (CBX7) interacted with ssRNA much more strongly than with the CD of mammalian HP1α protein (CBX5) (**Figure 2.5B**).

We then asked if this interaction occurred *in vivo*. Our collaborators in the Heard laboratory used IF to show that localization of CBX7-EGFP protein to the X_i diminished upon treating the cells with RNase (**Figure 2.5C**). Similarly, when I treated chromatin isolated from differentiated ESCs with RNase, I also saw loss of CBX7 association by immunoblot (**Figure 2.5D**); however, I did not see loss of the chromatin-associated WDR5 or HP1β proteins upon RNase treatment. Although

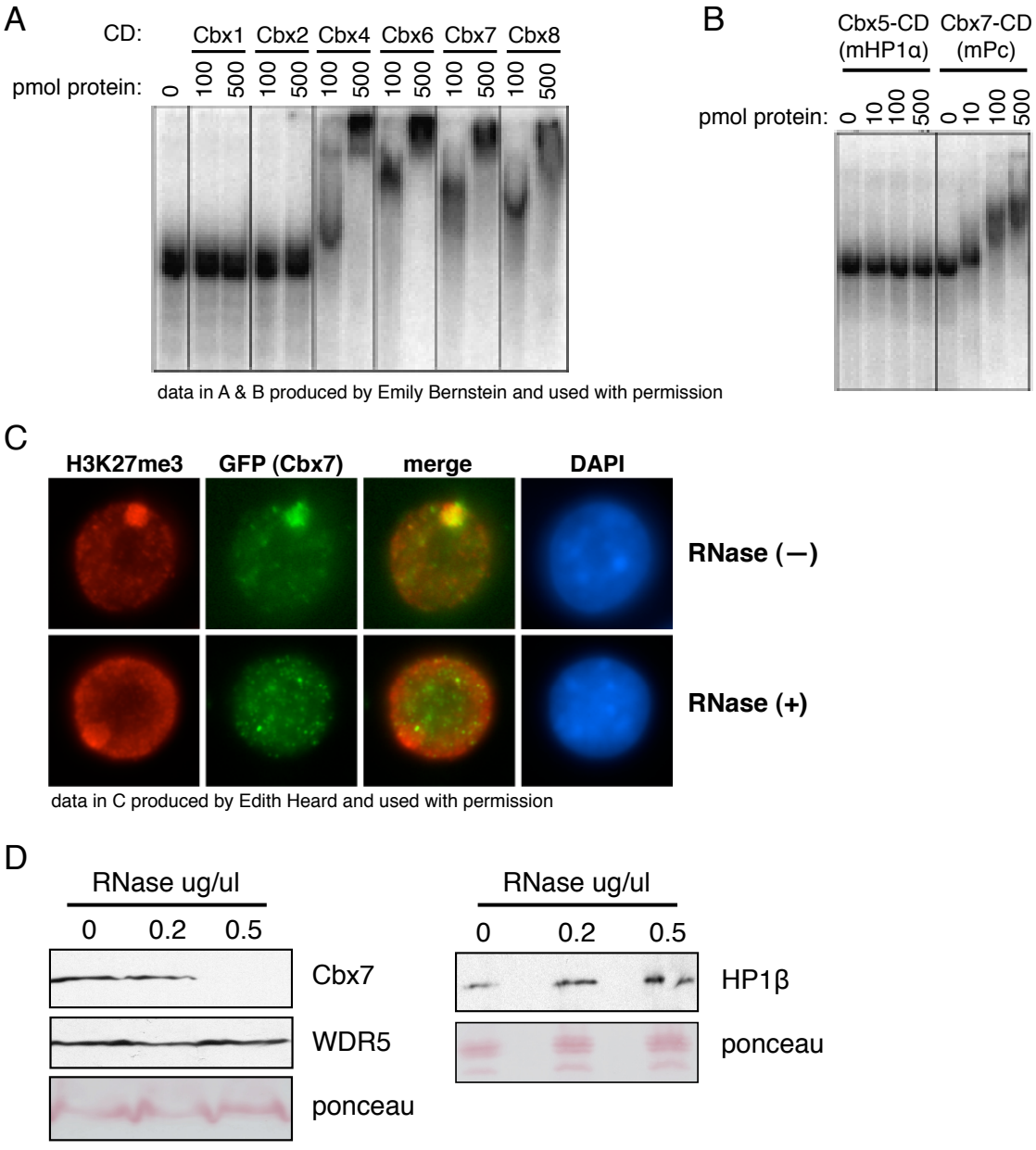


Figure 2.5: The CDs of some mammalian Pc/CBXs interact with RNA.

A. Recombinant CDs were tested for RNA binding by gel shift using a 500 nucleotide single-stranded (ss) RNA. CBX4, 6, 7, and 8 CDs interact with RNA, while CBX1 (mHP1β) and CBX2 do not.

B. The CBX5 (mouse HP1α) CD was compared to that of CBX7 for RNA binding; only the latter produces a shift. (Images in **A** and **B** produced by Emily Bernstein and used with permission.)

C. RNase-treatment of 3 day differentiated ES cells strongly diminished the accumulation of CBX7-EGFP signal on the X_i when examined by IF. The X_i is visualized by K27me3 staining (red) and the signal of the CBX7-EGFP fusion protein was enhanced using α-GFP antibody (green). DNA is stained with DAPI (blue). (Images produced by Edith Heard and used with permission.)

D. CBX7 is depleted from 6 day differentiated ES cell chromatin by RNase treatment (left), while WDR5 (left) and HP1β (right) are not affected. The PVDF membrane was stained with Ponceau to demonstrate that samples were loaded in roughly equal amounts.

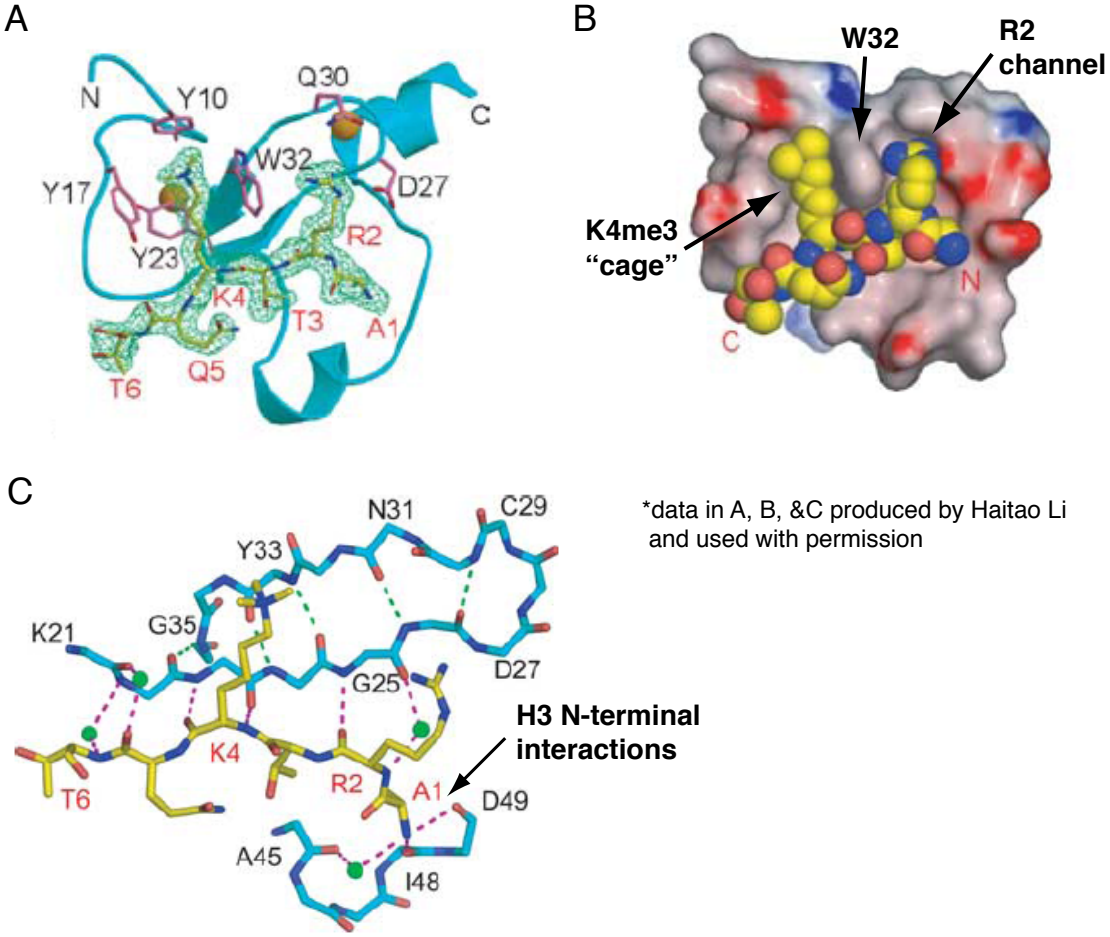
we did not see any sequence specificity in this interaction, it is interesting to hypothesize that the non-coding RNA Xist might play a direct role in this interaction *in vivo*.

A PHD Finger of the NURF Complex Binds H3K4me3 Via an Aromatic Cage

In collaboration with another post-doctoral fellow in our laboratory, Joanna Wysocka, I also studied the binding affinity of another methylated histone binding module, a PHD finger of the NURF complex component BPTF. After several studies (Ng et al., 2003; Santos-Rosa et al., 2002; Schneider et al., 2004) demonstrated that trimethylation of H3 at lysine 4 (H3K4me3) strongly correlates with transcriptional activation, many researchers in our field became interested in identifying H3K4me3 effector proteins in order to learn more about the mechanisms through which this modification affects transcription. Wysocka et al identified the PHD motif of BPTF as an H3K4me3 binding module after observing that the NURF chromatin-remodeling complex preferentially associated with H3K4me3 peptide over peptides that were either unmodified, mono- or dimethylated at H3K4 when all four peptides were incubated with nuclear extracts from mammalian cells (Wysocka et al., 2006). Since NURF is known to promote transcription (Georgel et al., 1997), this finding suggests that H3K4me3 may play a role in activating transcription through the stabilization of chromatin remodeling complexes (see Chapter 1 **and Figure 1.2B**). Moreover, the association was shown to be specific for H3K4me3 over other known trimethylated lysines in histone H3, specifically the repressive marks H3K9me3 and H3K27me3. In addition, only one of the two PHD fingers in human BPTF and one of the three in the homologous *Drosophila* protein NURF 301 (PHD2) was shown to bind H3K4me3 (Wysocka et al., 2006).

We then wanted to investigate the molecular basis for this interaction. We collaborated with the Patel laboratory at Memorial Sloan Kettering Cancer Center to solve the crystal structure of the BPTF H3K4me3-binding PHD-finger in both

the free and peptide-bound (H3 1-15 K4me3) states (Li et al., 2006). The crystal structure in complex with the H3K4me3 peptide (**Figure 2.6**) revealed two interesting features of this PHD finger: first, that the H3K4 trimethyl group is positioned within an aromatic cage composed of amino acids Y10, Y17, Y23 and W32 (which is reminiscent of the aromatic cage in the Pc/CBX chromodomain) and second, that H3 residue R2 and the H3K4 trimethyl group are positioned in adjacent surface



*data in A, B, &C produced by Haitao Li and used with permission

Figure 2.6: The molecular basis of the interaction between H3K4me3 and a PHD finger of the human NURF complex subunit BPTF.

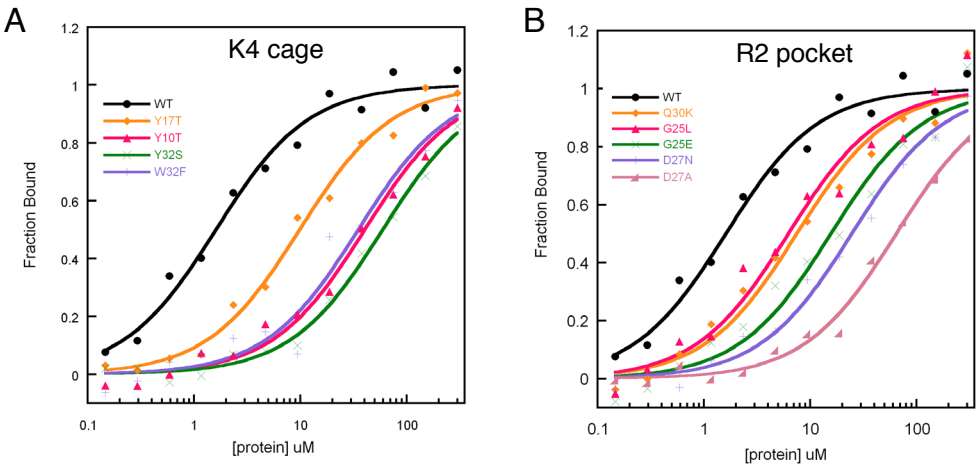
A. Ribbon representation of the crystal structure of the BPTF PHD finger + H3K4me3 peptide complex with the electron density highlighted for the bound H3K4me3 peptide (aa 1–15).

B. Electrostatic surface representation (red = negatively charged and blue = positively charged surface) of the crystal structure of the BPTF PHD finger + H3K4me3 peptide complex; the peptide is shown in a space filling representation. This view shows how K4me3 is positioned in a "cage" and arginine (R) 2 in an adjacent "channel", separated by tryptophan (W) 32 of BPTF.

C. Representation of the intermolecular backbone interactions between the A1-T6 segment of the H3K4me3 peptide and the PHD finger of BPTF. Note the interactions between the very N-terminus of the H3 peptide and BPTF. (Images in **A**-**C** produced by Haitao Li and used with permission.)

channels separated by residue W32 (see **Figure 2.6A and B**). These features not only demonstrate the molecular basis of the interaction between the PHD finger and the K4 trimethyl group, but also point to a possible explanation for the specificity of this interaction with H3K4me3 versus other methylated lysines, such as H3K9me3 or H3K27me3. The positioning of W32 to create two binding channels that fit H3R2 and H3K4me3 requires that they be separated by one amino acid, as they are in the H3 sequence surrounding K4 (i.e. R2-T3-K4) but not in the sequences surrounding H3K9 or H3K27 (i.e. R8-K9 or R26-K27). The structure also reveals that the amino terminus of the H3 1-15 peptide interacts with the PHD finger to anchor it, suggesting the possibility that steric hindrance caused by the longer stretch of residues amino to H3K9me3 and H3K27me3 (eight and 26 amino acids, versus 3 for H3K4) may also contribute to their lack of interaction with this PHD finger (**Figure 2.6C**).

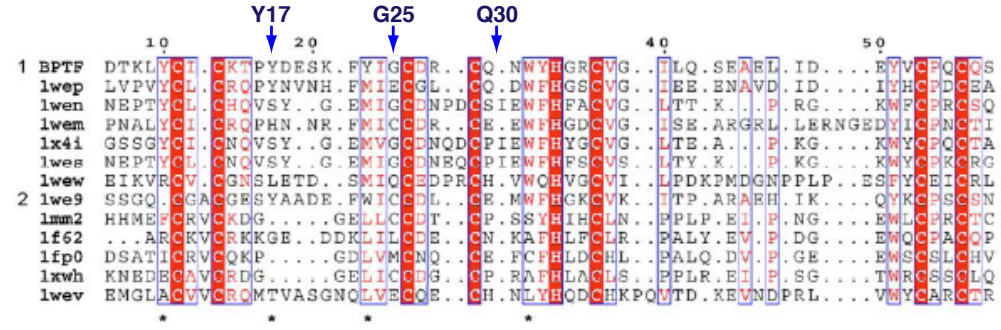
To test the influence of specific residues lining the aromatic cage and R2 channel, I measured the binding constants of the wild-type PHD finger of BPTF versus several mutants using fluorescence anisotropy. The Patel laboratory generated and purified point mutant proteins to each of the four aromatic residues in the cage (Y10, Y17, Y23, W32) and to three key residues in the R2 channel (G25, D27, Q30). As shown in **Figure 2.7**, wild-type BPTF interacted strongly to 1-15 H3K4me3 peptide with a K_d of 1.6±0.1μM while point mutations to both the aromatic cage (**Fibure 2.7A**) and R2 channel (**Figure 2.7B**) significantly decreased the binding affinity of these proteins (see **Figure 2.7C** for table of all K_ds). Of the cage mutants, Y17T affected binding the least (~6 fold decrease versus ~20-40 fold for the other three) and, interestingly, this residue is the least conserved of the four in comparison with PHD finger homologues (**Figure 2.7D**). Similarly, G25 and Q30 mutants affected binding somewhat less than mutations to D27, and these residues also vary between different PHD finger homologues.



C

Cage mutants	K_d H3K4me3 (μM)	Pocket mutants	K_d H3K4me3 (μM)
WT	**1.6 ± 0.1**	Q30K	7.5 ± 0.4
Y17T	10.0 ± 0.8	G25L	6.2 ± 0.3
Y10T	40.3 ± 4.5	G25E	15.8 ± 1.1
Y23S	59.7 ± 8.3	D27N	24.6 ± 0.6
W32F	34.4 ± 1.1	D27A	59.2 ± 3.4

D



data in D produced by Haitao Li and used with permission

Figure 2.7: Mutational analysis of H3K4me3-binding by the BPTF PHD finger.

A. Fluorescence polarization binding curves for FAM-labeled H3K4me3 peptide (aa 1-15) with BPTF PHD finger wild-type (WT) protein and proteins with K4me3-binding cage mutations.

B. Binding curve for peptide + WT protein and proteins with R2-binding pocket mutations.

C. Dissociation constants (K_D in μM) for the binding in **A** and **B**. Values represent mean ± SD for at least 3 independent experiments in all cases.

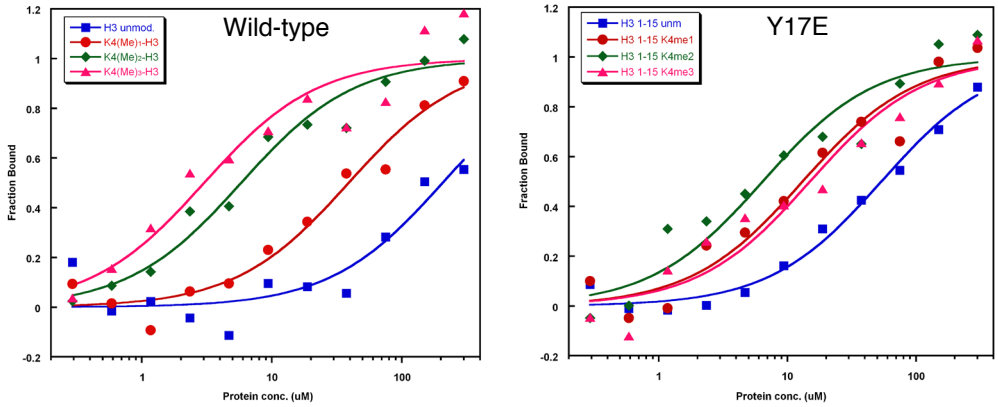
D. Alignment of the protein sequence of 12 different PHD fingers: 1wep=PHF8 [*Mus musculus*]; 1wen=ING4 [*Mus musculus*]; 1wem=DIDO1 (DATF-1) [*Mus musculus*]; 1x4i=ING3 [*Homo sapiens*]; 1wes=ING1-like [*Mus musculus*]; 1wew=AT5g60420 [*Arabidopsis thaliana*]; 1we9=Alfin-1-like [*Arabidopsis thaliana*]; 1mm2=MI2B/CHD4 2nd PHD domain [*Homo sapiens*]; 1f62=WSTF [*Homo sapiens*]; 1fp0=KAP-1 corepressor [*Homo sapiens*]; 1xwh=AIRE1 [*Homo sapiens*]; 1wev=PHF22 [*Mus musculus*]. There are two subsets (1 and 2) of PHD fingers in terms of CxC and CxxC spacing of the first zinc finger. * = residues that form the K4me3-binding cage. Note that most members in group 2 lack the aromatic cage residues of BPTF. Blue arrows point to those residues that are less well-conserved and cause less impact on binding affinity when mutated. (Alignment produced by Haitao Li and used with permission.)

A Co-Crystal Structure of L3MBTL1 and Dimethylated Lysine Peptide Provides Insight Into Aromatic Cage Binding Interactions

The above studies focused on two different protein motifs, the CD and one class of PHD fingers, that both bind trimethylated lysines via a cage of aromatic residues. Much less is known about protein motifs that bind mono- or dimethylated lysines. One of the few proteins that is know to interact preferentially with mono- and dimethylated lysines is the human Lethal (3) Malignant Brain Tumor-Like protein 1 (L3MBTL1) (Li et al., 2007). Its function suggested an interesting connection to chromatin in that it is a transcriptional repressor and is required for the progression of mitosis, perhaps through a role in chromosome compaction (Boccuni et al., 2003).

Our collaborators in the Patel Laboratory crystallized the structure of the three MBT repeats of this protein and showed that each repeat contains an aromatic cage motif, much like that of the CD and PHD finger (Wang et al., 2003). Further crystallization and quantitative binding studies confirmed that pocket 2 of L3MBTL1 binds mono- and dimethylated lysine residues strongly and preferentially, although, interestingly, without strong sequence specificity (Li et al., 2007). Among those peptides tested, L3MBTL1 pocket 2 bound mono- and dimethylated H1K26, H3K4, H3K9, H3K27, H3K36, and H4K20. We then compared the co-crystal structures of L3MBTL1 and BPTF and the H3K4me peptides (H3K4me1 and me2 peptides for L3MBT and H3K4me3 for BPTF) in order to learn more about the molecular differences that might potentially allow the former to prefer mono- and dimethylated lysines and the latter trimethylated H3K4. Our structural collaborators in the Patel laboratory concluded that aromatic cage of L3MBTL1 pocket 2 is deep and narrow, and therefore restricted to interact with the smaller mono- and dimethyl groups, while that of BPTF-PHD is more of a shallow cleft and can accommodate the larger trimethyl group.

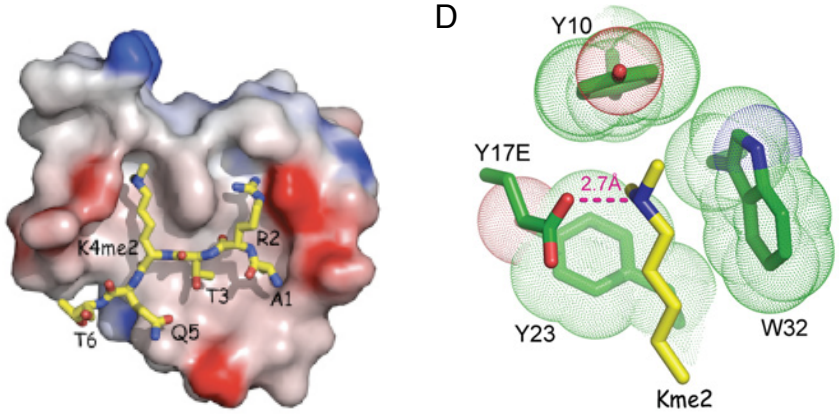
A



B

Peptide (H3 1-15)	K_d μM			
	WT	Y17E	Y17Q	Y17D
unmod	231±144	53±4	144±45	117±34
K4me1	46±33	13±1	46±5	29±9
K4me2	6±3	7±1	19±1	40±2
K4me3	3±0.3	15±3	27±9	25±7

C D



data in C & D produced by Haitao Li and used with permission

Figure 2.8: The Y17E mutant of a PHD finger of BPTF has increased affinity for H3K4me2 compared to the wild-type protein.

A. Fluorescence polarization binding curves for fluorescein-labeled unmodified, H3K4me1, 2 and 3 peptides (aa 1-15) with the WT BPTF PHD finger protein (left) and the Y17E mutant (right).

B. Dissociation constants (K_D in μM) for the unmodified, H3K4me1, 2 and 3 peptides (aa 1-15) with the WT BPTF PHD finger and the Y17E, Y17Q and Y17D mutants. Values represent mean ± SD for at least 2 independent experiments in all cases.

C. Electrostatic surface representation (red = negatively charged and blue = positively charged surface) of the crystal structure of the BPTF PHD finger + H3K4me2 peptide (aa 1-9) complex. This view shows the surface groove recognition mode of binding and complex formation.

D. A closer view of the dimethyl-lysine sidechain of the H3K4me2 peptide (aa 1-9) inserted into the engineered Y17E BPTF PHD finger mutant. The aromatic residues of the K4me-binding cage are shown in the dotted van der Waals representation. The carboxylate of the mutated residue E17 forms a direct hydrogen bond with the dimethylammonium group of K4me2 when positioned in the correct orientation. (Images in **C** and **D** produced by Haitao Li and used with permission.)

To test this theory, we mutated the PHD finger of BPTF to see if we could engineer a pocket more similar to that of L3MBTL1 and subsequently change its binding preference from tri- to dimethylated lysine. One of the four residues making up the aromatic cage of the BPTF-PHD finger, Y17, was mutated to glutamic acid, glutamine, and aspartic acid to assay the effects of changes in size and charge at this position. I then compared the binding affinity of these mutant proteins to wild-type BPTF-PHD finger using fluorescence anisotropy. As shown in **Figure 2.8A**, mutating Y17 to E successfully reversed the preference of the mutated PHD finger from trimethylated lysine to dimethylated lysine. This mutation also improved the binding affinity of this PHD finger for monomethylated lysine (**Figure 2.8B**). Mutations of Y17 to D or Q reduced the overall binding of BPTF-PHD for all three methylation states (**Figure 2.8B**). The Patel laboratory then crystallized the Y17E mutant in complex with dimethylated lysine peptide to show that this difference in binding was not due to an artificial change in protein folding (**Figure 2.8C**). They showed that the alignment of the 1-9 H3K4me2 peptide was almost identical to that shown in our earlier work (see above) with an H3K4me3 peptide, with only a single critical recognition difference in the contacts between K4me2 and the glutamic acid residue at position 17 (**Figure 2.8D**).

Chapter 2 Discussion

The idea that proteins bind to histone modifications and lead to biological outputs is strongly supported (Bannister et al., 2001; Fischle et al., 2005; Wysocka et al., 2005; Wysocka et al., 2006). In collaboration with my colleagues in the Allis Laboratory and with the Heard and Patel Laboratories, I performed studies that support the idea that the aromatic cage motif that is a recurrent and crucial structural feature of many methylated-histone binding proteins. Our studies show that even subtle differences in sequence between similar domains, whether natural (e.g. CBX4-CD vs CBX7-CD) or engineered (e.g. BPTF-PHD wild-type versus Y17E), can impact significantly the binding affinity of these "effector" proteins for their histone substrates, which could therefore affect the functional readout of histone methylation.

Fischle et al demonstrated *in vitro* that the chromodomain (CD) of chromatin-associating protein dmHP1 binds strongly to H3K9me3 peptide versus a non-modified control, and that of dmPc preferentially binds H3K27me3 peptide (**Figure 2.1A**) (Fischle et al., 2003). However, as described above, although the CDs of mammalian Pc-like proteins (CBX2, 4, 6, 7, 8) are highly similar to that of *Drosophila* Pc (**Figure 2.1B**) and bind H3K27me3 peptides, some also bind H3K9me and one even prefers K9me (**see Figures 2.1C and 2.2**). We also see this differential binding *in vivo* in the context of the mammalian inactive X chromosome (**Figure 2.3**). Importantly, the differential binding preferences of these proteins suggests that such binding may translate to distinct biological readouts. Nevertheless, it is still unclear how such a highly conserved domain binds these marks differentially.

One possible approach to solving this question is to perform structural studies on the CD of each mammalian CBX in complex, such as those described above in collaboration with the Patel laboratory. As we learned from the structural studies of BPTF and L3MBTL1, certain residues are often critical in both the recognition

of backbone sequence and the accommodation of the methyl group (**Figures 2.6, 2.7 and 2.8**). Crystal structures and/or NMR studies of the five mammalian CBX CDs could reveal which residues differ in their contacts with the H3 peptide backbone sequence and/or the methyl group at K9 or K27, which would then allow for mutagenesis studies to confirm such observations and/or to test hypotheses as to which residues might be mutated so that, for example, the CBX4 CD now prefersH3K27me3 rather than H3K9me3.

There are also many questions remaining about the function of the CBX CDs in the context of the full proteins *in vivo*. As described above, although both CBX7 and H3K27me3 are present throughout the differentiation of ESCs, the mammalian Pc protein CBX7 only associates with chromatin upon ESC differentiation (**Figure 2.4)**. On the other hand, HP1, which also binds silent chromatin marks, showed no difference in global chromatin association in undifferentiated versus differentiating ESCs. Together, these results strongly suggest that there is a mechanism, or several concerted mechanisms, that specifically and developmentally regulate the chromatin-association of CBX7. Moreover, they suggest that this association cannot be mediated by H3K27me3-binding alone, which is supported by the findings of others as well (Schoeftner et al., 2006; Vincenz and Kerppola, 2008).

Several mechanisms could lead to this regulated chromatin-association: the post-translational modification of other histones and/or with marks besides H3K27me3, differential post-translational modification of the CBX7 protein itself, or the formation of different biochemical complexes and/or the association with transient complex members at different stages of differentiation. One hypothesis as to why CBX7 is present in the nucleus but remains unassociated with chromatin in the early days of differentiation is that the protein itself is differentially glycosylated and phorphorylated post-translation. Using algorithms I found on several website, I discovered that CBX7 has several O-linked glycosylation and phosphorylation sites. These modifications are often thought to regulate protein function in a “ying/

yang" dualism, in which the glycosylation of the protein at a serine or threonine site has one effect (i.e. prevents the protein from binding chromatin) and the phosphorylation of that same site has the opposite effect (i.e. allows the protein to bind chromatin). It would be interesting to test whether CBX7 is differentially glycosylated during differentiation. Possible experiments include simple blots assays with biotin labeled lectins and streptavidin HRP, as well as using wheat germ agglutin conjugated agarose beads to enrich for glyscolylated proteins in nuclear and chromatin extracts from different time points of differentiation. In both types of experiments the available CBX7 antibody could be used to confirm whether a lectin tagged/purified band corresponds to CBX7 protein.

Although mammalian PcG protein targets have been studied extensively in recent years (Boyer et al., 2006; Bracken et al., 2006; Bracken et al., 2007; Lee et al., 2006), some very basic, yet critical, questions still remain unanswered. First, what is the minimal mechanism(s) that is sufficient for the recruitment of mammalian Polycomb Repressive Complex 1 (PRC1) group proteins (which include CBX) to chromatin? The fly genome contains conserved Pc binding elements (or Polycomb Repressive Elements, PREs), yet their counterparts have yet to be defined in mammalian cells, if they exist at all (Otte and Kwaks, 2003). The work described above suggests the CD may play an important role in recruitment by binding H3K27me3. Our data also suggests that an interaction between the CBX-CDs and RNA may be involved in their recruitment to chromatin. Moreover, the ability of particular CDs to bind both methylated histone tails and RNA suggests that a cooperative binding mechanism may mediate their enrichment in chromatin. Functional and structural studies will be required to determine the nature of this potential synergy; Assistant Professor Emily Bernstein is currently conducting such studies in her laboratory at Mount Sinai School of Medicine. Nevertheless, the exact mechanism through which the CD enables PRC1 recruitment and/or stabilization on chromatin is still not known.

Second, what is the mechanism of silencing through the recruitment of repressive binding proteins such as Pc? There is evidence for PcG protein dimerization, including *in vitro* data for the self-interaction of dmPc (Min et al., 2003) and the Xenopus homologue M33 (Reijnen et al., 1995), and both *in vitro* and *in vivo* data showing that Mel-18 dimerization is regulated by its phosphorylation (Fujisaki et al., 2003). Such interactions may function to cause heterochromatin compaction and therefore silencing. However, studies in ESCs showing that H3K27 methylation via PcG proteins is dominant over H3K4 methylation at bivalent domain loci (Azuara et al., 2006) suggests that complete heterochromatin compaction is not necessary for silencing. The work described above, as well as that of others (Zhao et al., 2008), suggests that PcG protein interactions with non-coding RNAs may play a role, particularly in the case of Xist and the mammalian inactive X chromosome. Importantly, PcG proteins also include Ring1A and B, which have ubiquitinating activity (de Napoles et al., 2004). It may be that mammalian Pc proteins (CBXs) employ different modes of silencing by varying their interactions with other proteins; data from the Reinberg lab shows that this type of complex-swapping regulation occurs during differentiation amongst the proteins of the PRC2 complex (Kuzmichev et al., 2005). Moreover, very recent work suggests that PcG proteins (specifically Polycomb Repressive Compex 1 (PRC1) proteins, which include Pc itself) can remain bound to chromatin during DNA replication (Francis et al., 2009). This implies that PcG proteins are not only critical to the regulation of gene silencing, but also to the inheritance of this silent status through cell division.

The studies described above build upon the work of many other groups and support the idea that interactions between covalent modifications on histone proteins and non-histone chromatin proteins can have significant biological consequences. Moreover, they emphasize the fact that conservation between

primary protein sequences does not provide complete information about the *functional* conservation of domains such as the methyl-binding CD. Structural and mutagenesis studies, such as those described for the methyl-binding PHD finger and MBT repeats, can pinpoint features that are critical to specific protein-protein interactions and potentially help to create testable models and hypotheses about how these domains mediate biological function.

The studies described in this Chapter focus on the specific features of various non-histone chromatin proteins that mediate their binding to particular histone modifications. However, most of these experiments were done *in vitro* with histone peptides that are a simplified representation of the canonical histone H3. In the next few chapters, I will describe how the H3 substrate actually changes significantly during the process of mouse ESC differentiation. Moreover, in Chapter 4 I will describe how one of these changes, in particular, impacts the affinity of the CBX7-CD for H3K27me3.

Chapter 3: Analysis of global histone H3 profiles across ESC differentiation

Summary

As described in Chapter 2, I began using a female ESC model in order to study the recruitment of mammalian Polycomb proteins to chromatin during differentiation. In this chapter, I will describe two studies in which I analyzed the global profile of histone H3 in these ESCs during differentiation. In the first study, post-doctoral fellow Sandra Hake and I used biochemical methods to separate and measure the three histone H3 isoforms in mouse ESCs at different stages of differentiation. In the second, I performed chromatin immunoprecipitation (ChIP) with an antibody to the well-characterized histone modification H3K4 trimethylation (H3K4me3) and then collaborated with both fellow student Aaron Goldberg to sequence the associated DNA and Assistant Professor Deyou Zheng at Albert Einstein College of Medicine to analyze those sequences. In both studies, our aim was to see how H3 isoforms and modifications known to correlate with gene activation change during the process of ESC differentiation. The results of these studies support both the findings of previous reports (Azuara et al., 2006; Bernstein et al., 2006; Lennox and Cohen, 1988; Pina and Suau, 1987) and, importantly, the validity of this ESC differentiation model.

Results

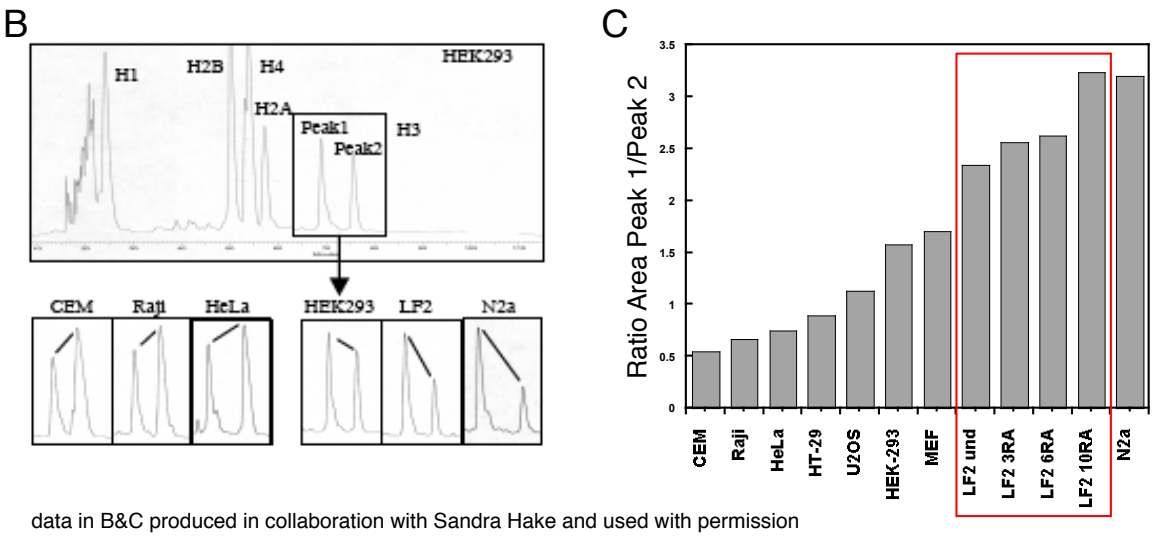
Incorporation of H3 Isoforms Into Chromatin During ESC Differentiation

Histone proteins are among the most conserved proteins in the proteome. Histone H3 is particularly well conserved, although its diversity is increased by the expression of more than one isoform in multicellular organisms. Mammalian cells have been shown to express three "core" histone isoforms: H3.1, H3.2 and H3.3 (**Figure 3.1A**). H3.1 and H3.2 are incorporated into chromatin in a replication-dependent manner (RD); H3.3, on the other hand, can be incorporated into chromatin at any point in the cell cycle, including both S-phase and outside of replication (replication-independent, RI; **see Figure 1.5B**). Although these isoforms differ by at most five amino acids (i.e. mammalian H3.1 compared to H3.3), they have been shown to correlate with functionally distinct regions of the genome. In *Drosophila*, RI H3.3 was shown to be enriched over the RD H3 isoform in regions of chromatin undergoing active transcription (Ahmad and Henikoff, 2002). Importantly, the incorporation of RI H3.3 into chromatin outside of S-phase was abolished when the four differing amino acids between the two isoforms were mutated to those of the RD H3. In addition, other studies have shown that these isoforms associate with distinct protein complexes (Tagami et al., 2004) that serve as chaperones for the deposition of these H3 isoforms into chromatin.

My colleague Sandra Hake and I were interested in analyzing the relative amounts of histone H3 isoforms in the chromatin of mammalian cells. Since there are currently no antibodies available that can distinguish between the three H3 isoforms, she chose to use two different biochemical methods to separate and measure the relative amounts of these proteins in mammalian cells. First, she used reverse-phase high-pressure liquid chromatography (RP-HPLC) to separate total histone extracts (**Figure 3.1B**). Although many previous studies had shown

A

```
                    31   87   89 90                  96
Hs/Mm  H3.1:  ...-A-...-S- - -V-M- - - - - - - - - - C....
Hs/Mm  H3.2:  ...-A-...-S- - -V-M- - - - - - - - - - S....
Hs/Mm  H3.3:  ...-S-...-A- - -I-G- - - - - - - - - - S....
```



data in B&C produced in collaboration with Sandra Hake and used with permission

Figure 3.1: Mammalian cell lines differ in their H3 protein composition.

A. Sequence alignment of the three histone H3 variants: H3.1, H3.2 and H3.3. H3.1 and H3.2 deposition occurs during S-phase and is therefore dependent on replication, while H3.3 deposition occurs both during and outside of S-phase and is therefore independent of replication. Differences in amino acid sequence are shown in colored font (black = "canonical" H3.1 sequence, colors = differences found in H3.2 and H3.3); identical amino acids are shown as dashes (-).

B. Top: a typical Reverse Phase-High Pressure Liquid Chromatography (RP-HPLC) profile of HEK293 histones. Peaks that correspond to individual histones are labeled. Histone H3 elutes as two peaks (boxed): peak1 and peak2. Bottom: RP-HPLC profiles of histone H3, peak1 and peak2, from six representative mammalian cell lines.

C. Ratios of RP-HPLC peak areas (peak1/peak2) from H3 samples isolated from nine different mammalian cell lines, including data for undifferentiated mouse ESCs (LF2) and ESCs treated with Retinoic Acid (RA) for 3, 6, and 10 days (red box).

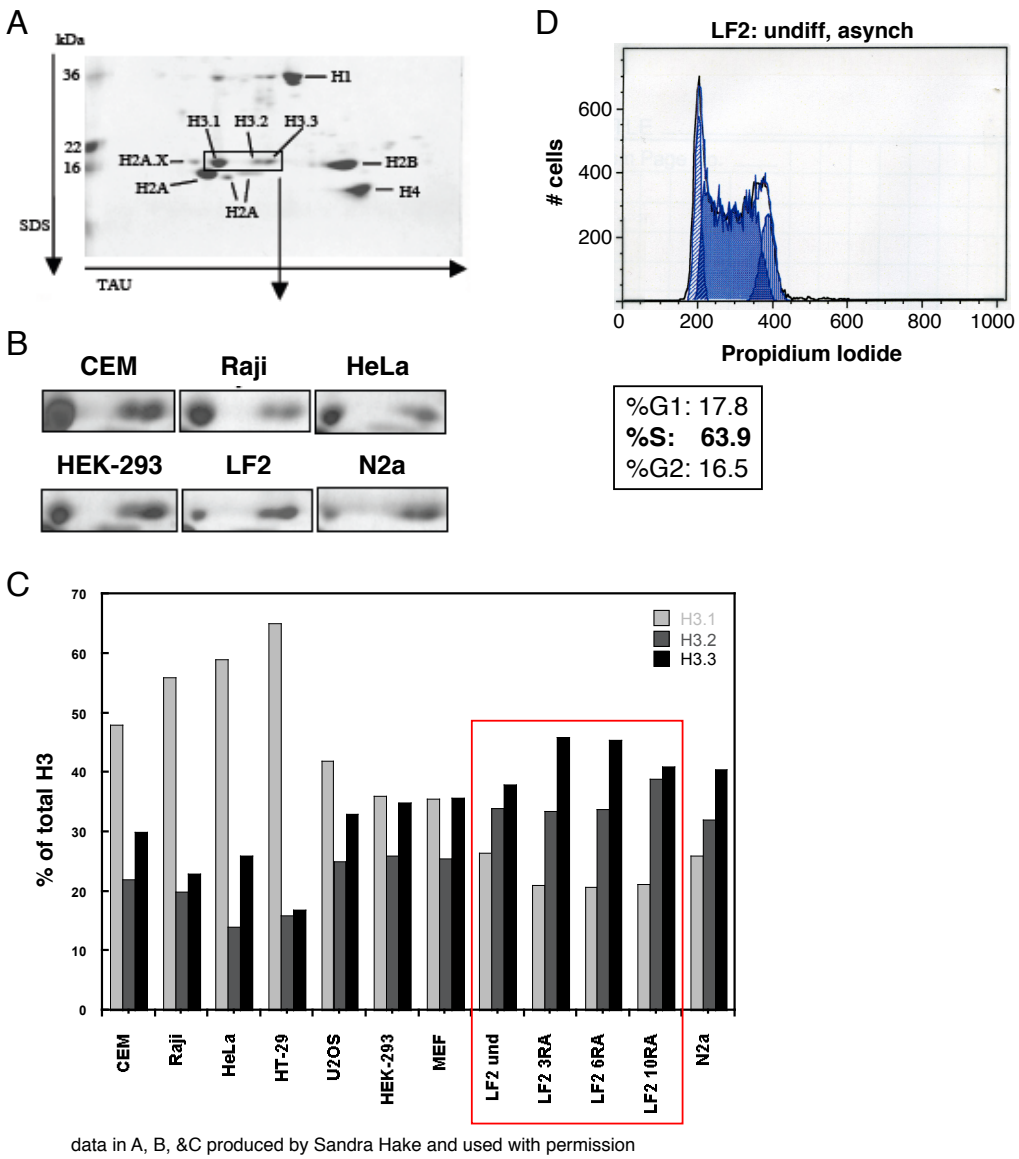
---

that H3 elutes in two separate peaks (i.e. "peak 1" and "peak 2"), Sandra noted that the ratio of the areas of these H3 peaks differed between different mammalian cell lines. Mass spectrometry analysis of these peaks by Benjamin Garcia of the Hunt Laboratory (UVa) confirmed that peak 1 consisted of H3.2+H3.3 and peak 2 of H3.1. Sandra therefore hypothesized that the change in peak area ratio reflected a change in H3 isoform deposition into chromatin.

We then wondered how this ratio might change in ESCs as they differentiated. Using the same retinoic acid (RA) differentiation method described in Chapter 2, I differentiated female mouse ESCs (LF2 line) with RA for 3, 6 and 10 days. I then extracted the histones from these cells, separated them by RP-HPLC, and calculated the peak area ratios between H3 peaks 1 and 2. As shown in **Figure 3.1C**, the ratio of peak1/peak2 increases as ESCs differentiate (see data boxed in red). We also noted that, interestingly, both ESCs and the neuroblastoma cell line N2a had much higher peak1/peak2 ratios than other cell types tested.

Since the differences in these ratios could be due to an increase in H3.2 or H3.3 or a decrease in H3.1, we then wanted to develop a way to separate all three isoforms. We turned to a 2D gel electrophoresis system, in which the first dimension was separation by Triton-Acid-Urea (TAU) gel electrophoresis and the second was separation by standard SDS-PAGE gel electrophoresis. As shown in **Figure 3.2A**, this method allowed us to separate all three histone H3 isoforms. We confirmed that the boxed bands (**Figure 3.2A**) were indeed H3.1, H3.2 and H3.3 by mass spectrometry (performed by Benjamin Garcia in the Hunt Laboratory). We then used this 2D gel system to separate total histones extracted from each cell type (**Figure 3.2B**) and used Image Gauge software to calculate the relative amount of each H3 isoform for each cell type (**Figure 3.2C**).

This quantification revealed several interesting differences, both between the stem-like cells (LF2 and N2a) and other cell lines and within the different stages of ESC differentiation. First, stem-like cells appear to incorporate a relatively low amount of H3.1 into their chromatin versus H3.2 or H3.3 as compared to other cell types. Cells such as Raji and HeLa have a much higher relative amount of H3.1 than H3.2 or H3.3, whereas the opposite is true of LF2 or N2a cells. This supports the data in **Figure 3.1** showing that the stem-like cells have greater peak1/peak2 ratios. Second, H3.3 appears to increase in its relative amount upon differentiation



data in A, B, &C produced by Sandra Hake and used with permission

Figure 3.2: Mouse ESCs have an increased amount of the variant H3.3 as compared with more differentiated cell lines.

A. An example of total acid-extracted histones separated by 2-dimension (2D) gel electrophoresis: the first dimension separation is by Triton Acid Urea (TAU) gel electrophoresis and the second by SDS/PAGE. The gel is then stained with Coomassie blue. Mammalian histone H3 variants separate into three different spots (box): H3.1, H3.2, and H3.3, from left to right.

B. The region containing H3.1, H3.2, and H3.3 (boxed in **A**) of 2D TAU gel analysis (as in **A**) of total acid-extracted histones from representative mammalian cell lines (see **Figure 3.1C**).

C. Quantification of H3.1, H3.2, and H3.3 proteins in 2D TAU gels (shown in **B**). Total H3.1, H3.2, and H3.3 protein was set as 100% and the distribution of H3.1 protein (light gray bars), H3.2 (dark gray bars), and H3.3 (black bars) is shown as % of total H3. Data for undifferentiated mouse ESCs (LF2) and ESCs treated with Retinoic Acid (RA) for 3, 6, and 10 days are boxed in red.

D. DNA histogram of asynchronously growing undifferentiated ESCs stained with propidium iodide and analyzed by flow cytometry. Data was analyzed using FloJo FACS software.

of ESCs, although the latest point of differentiation (10 days post RA induction) shows a relative increase in H3.2 as well. This second finding is interesting in that it both agrees with and raises interesting questions about some hypotheses of differential H3 isoform deposition.

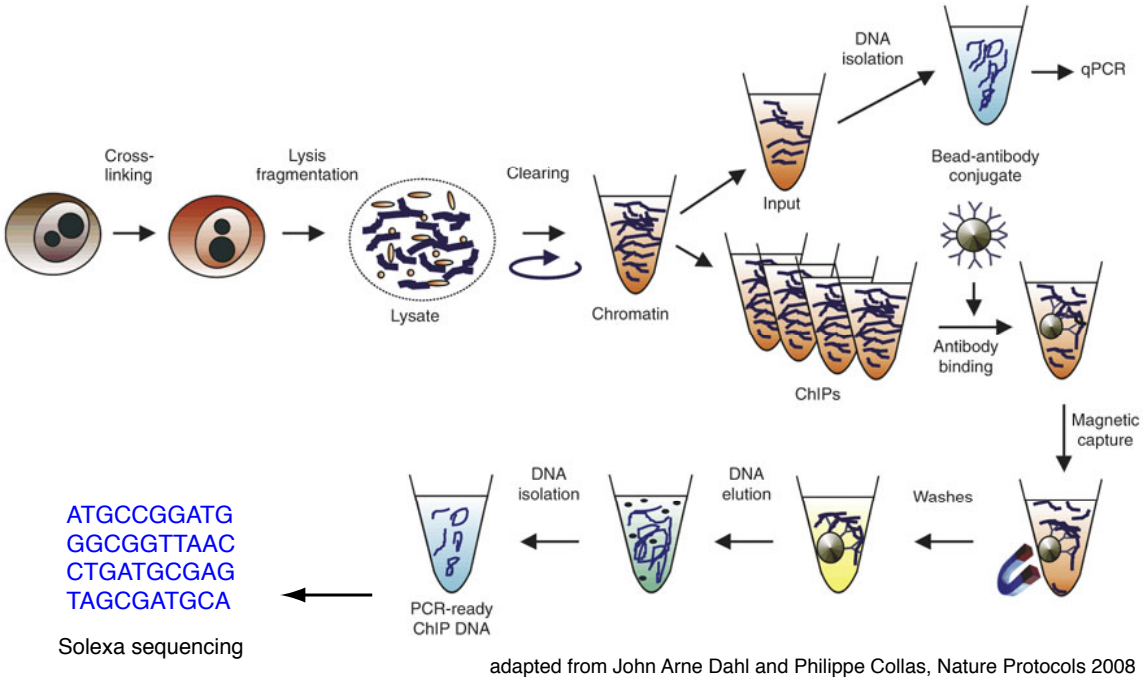
Since H3.3 is deposited into chromatin in a replication-independent manner, it has been hypothesized that its relative amount should increase upon differentiation into non-dividing cells such as neuronal cells (Lennox and Cohen, 1988; Pina and Suau, 1987). Indeed, rat tissue types such as heart and brain that are enriched in post-mitotic cells were found to have higher levels of H3.3 protein, when analyzed by RP-HPLC (Garcia et al., 2008). Our findings support this hypothesis in that both ESCs differentiated toward a neuronal lineage with RA and the stem-like neuroblastoma cell line N2a have increased relative amounts of H3.3. At the same time, given that stem-like cells cycle and proliferate very rapidly compared to other cell lines and spend more time in S-phase (as shown by FACS profile in **Figure 3.2D**, ~64% of undifferentiated ESCs are in S-phase as compared to ~10-30% in most other asynchronously grown cell lines), it is interesting that these cells have so much H3.3 incorporated into chromatin. This suggests that there must be an active remodeling and/or deposition mechanism to maintain this elevated ratio of RI H3 to RD H3, or that these cells express relatively high levels of H3.3 protein, which could be incorporated during as well as outside of S-phase.

Changes in H3K4me3 Localization During ESC Differentiation

In addition to the incorporation of different isoforms of the histone H3 protein, variation is also introduced into the chromatin template by enzymes that add (and remove) post-translational covalent modifications to specific residues of H3. One well-characterized modification is H3K4me3, which has been shown to correlate with the transcription start site of genes (Barski et al., 2007). This correlation occurs predominantly at gene loci that are being actively transcribed,

although studies have also shown that H3K4me3 also localizes at lower levels to the transcription start site of genes that are not being expressed (Guenther et al., 2007). One interesting example of this is found in ESCs, in which Bernstein and colleagues have shown that H3K4me3 co-occupies many of the same loci as H3K27me3, a modification that has been correlated with transcriptional silencing (Bernstein et al., 2006). They also demonstrate that this duality of modifications is mostly "resolved" upon differentiation, so that only one modification (H3K4me3 or H3K27me3) localizes to a particular locus, and that this resolution restores the correlation between H3K4me3 and active transcription. These studies and others (Azuara et al., 2006) suggest the possibility that H3K4me3 correlates both with genes that are being actively transcribed as well as those that are simply "poised" for transcription, i.e. transcription can quickly proceed upon removal of the dominant silencing mark H3K27me3.

This interesting finding prompted me to investigate the localization of H3K4me3 in differentiating ESCs using the method I typically employ (see Chapter 6: Materials and Methods). Chromatin immunoprecipitation (ChIP) is a method used to map the loci at which a particular histone modification is enriched or the target loci of a given non-histone chromatin protein. The schematic shown in **Figure 3.3** depicts the steps of a typical ChIP experiment. Using a well-characterized and highly specific antibody to H3K4me3, I performed ChIP on the cross-linked chromatin of both undifferentiated ESCs and ESCs differentiated for 3 days with RA. I then assayed for the presence of specific DNA sequences by Q-PCR with Sybergreen to test the success and specificity of the ChIP. I first tested basic positive and negative controls (**Figure 3.4A**). Since the actin gene is highly expressed in both undifferentiated and differentiated ESCs, and therefore presumably marked by H3K4me3, I assayed for the presence of actin DNA in these ChIP samples. H3K4me3 was indeed associated with actin, with an enrichment of



adapted from John Arne Dahl and Philippe Collas, Nature Protocols 2008

Figure 3.3: Chromatin ImmunoPrecipitation (ChIP) is a method for determining the genomic targets of chromatin proteins.

A schematic of the Chromatin ImmunoPrecipitation (ChIP) method for isolating gemonic sequences associated with a given chromatin protein. The DNA is first fixed to nearby proteins by cross-linking with formaldehyde, and then sheared into ~500-1000 base pair segments through sonication to form a lysate. Cells that have not sheared well and other non-soluble debris is cleared by centrifugation and the resulting supernatant is divided into "input" and "ChIP" samples. Magnetic beads coated in the chosen antibody are then incubated with the ChIP samples to allow the binding of the antibody with its specific epitope(s). The beads are then washed of any non-specific interactions. The remaining chromatin fragments are then eluted from the beads and treated (along with the input sample) to reverse the cross-linking between DNA and protein. The DNA is then isolated from input and ChIP samples and ready for quantitative PCR and/or sequencing.

7.8% and 11.1% over input DNA for undifferentiated and 3 days +RA differentiated cells, respectively. The 3 days +RA differentiated cells showed higher enrichment of H3K4me3 at the actin gene in this particular experiment, but only ~1.4 fold. Conversely, Q-PCR for a DNA sequence in an intergenic region of the mouse genome did not show any significant enrichment over input (< 0.5%).

I was then interested in testing genes associated with either pluripotency (markers of undifferentiated ESCs) or a particular lineage (markers of differentiation). As shown in **Figure 3.4B**, representative markers of both pluripotency and lineage behaved as expected: Oct3/4, a marker of pluripotency, showed greater enrichment

Figure 3.4: H3K4me3 is differentially enriched at stem cell marker Oct3/4 and differentiation marker Sox17 in undifferentiated and 3 days +RA differentiating ESCs, respectively.
A. H3K4me3 ChIP data for both negative (an intergenic region) and positive (β-actin) control ChIP targets as assayed by RT-PCR. Input and H3K4me3 antibody ChIP eluates from both undifferentiated ESC (black) and 3 days +RA differentiating ESC (grey) samples were assayed for the presence of intergenic region and β-actin DNA by RT-PCR with SYBRgreen. Signal for each DNA primer pair (intergenic region or β-actin) with each sample (undifferentiated ESCs or 3 days +RA differentiated ESCs) was normalized to the corresponding input signal and plotted as shown.
B. H3K4me3 ChIP data for both a stem cell marker (Oct3/4) and a marker of differentiation (Sox17) as assayed by RT-PCR with SYBRgreen.
C. Gene expression (mRNA) data for both a stem cell marker (Oct3/4) and a marker of differentiation (Sox17) as assayed by RT-PCR with SYBRgreen.
D. H3K4me3 ChIP data for both a stem cell marker (Oct3/4) and a marker of differentiation (Sox17) as assayed by Solexa sequencing. BED file data generated by The Rockefeller University Genomic Resource Center was uploaded to the UCSC Genome Browser, which then mapped the sequenced DNA to the mouse genome. Each box represents a representative gene locus (Oct3/4 or Sox17). Sequence tags that map to a given region are represented as red and blue vertical dashes. The location of the selected gene is labeled at the top of each box; its exon-intron organization is labeled at the bottom.

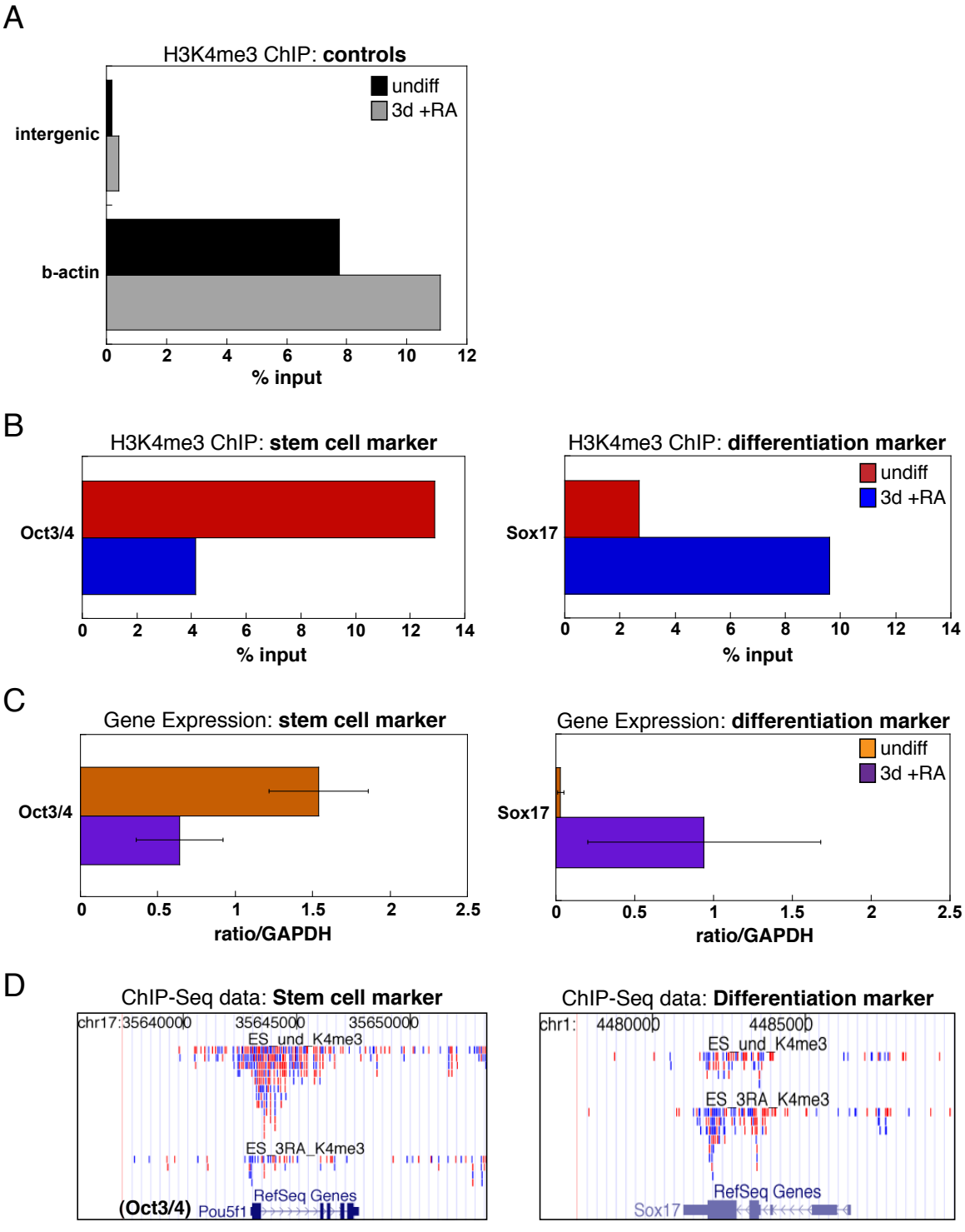


Figure 3.4: H3K4me3 is differentially enriched at stem cell marker Oct3/4 and differentiation marker Sox17 in undifferentiated and 3 days +RA differentiating ESCs, respectively.

for H3K4me3 in undifferentiated cells (12.9%) than in 3 days +RA differentiated cells (4.1%, ~3.1 fold change); Sox17, a marker of endodermal differentiation, showed greater enrichment for H3K4me3 in 3 days +RA differentiated cells (9.6%) than in undifferentiated cells (2.7%, ~3.6 fold change. Importantly, expression of both the Oct3/4 and Sox17 genes, as measured by Q-PCR of mRNA levels, correlated with the enrichment of H3K4me3 at these loci (**Figure 3.4C**): Oct3/4 showed greater expression in undifferentiated ESCs as compared to 3 days +RA differentiated ESCs, while Sox17 showed greater expression in the differentiated cells.

After establishing that ChIP with H3K4me3 antibody produced the expected results for the above control genes in undifferentiated and 3 days +RA differentiated ESCs, I wanted to perform a more comprehensive, less biased analysis of the genes that associated with H3K4me3 these samples. I therefore collaborated with a fellow graduate student in our laboratory, Aaron Goldberg, who has optimized a method for ChIP followed by sequencing with Solexa sequencing technology (ChIP-seq). After performing the above ChIP, the samples were prepared for sequencing by adding specific adaptors to the precipitated DNA and amplifying the sequences by PCR. We then submitted these samples to The Rockefeller University Genomic Resource Center for sequencing. In order to assess whether the sequencing results agreed with those generated by RT-PCR, I mapped the Solexa sequences to the mouse genome using the University of California at Southern California (UCSC) Genome Browser. As shown in **Figure 3.4D**, the Solexa sequencing data correlated well with the traditional ChIP data in **Figure 3.4B**. At the locus of stem cell marker Pou5f1 (Oct3/4, **left**), many more sequencing tags map there in the undifferentiated ESC sample (top line) than to the 3 days +RA differentiating sample (bottom line). In contrast, many more sequencing tags map to the 3 days +RA differentiating sample (bottom line) than to the undifferentiated ESC sample (top line) at the locus of differentiation marker Sox17 (**right**).

We then collaborated with Assistant Professor Deyou Zheng at Albert Einstein College of Medicine to perform a global analysis of the data. Dr. Zheng generated lists of genes that were differentially enriched in H3K4me3, i.e. genes that are significantly marked by H3K4me3 in undifferentiated and those that are enriched in the 3 days +RA sample. The lists included > 300 genes marked by H3K4me3 in the undifferentiated sample and >500 genes that show enrichment in the 3 days +RA sample (**see Appendix** for complete list). To check the validity of this analysis, I used the UCSC Genome Browser to map the raw sequence data and check representative genes. Included in the list of genes marked by H3K4me3 in undifferentiated ESCs was the stem cell marker Nanog. As shown in the snapshot of the Nanog locus in the UCSC Genome Browser (**Figure 3.5A, panel a**), our undifferentiated sample (top line) indeed contained many more sequencing tags mapping to the Nanog locus than did the 3 days +RA differentiating sample (bottom line). This result correlated, as expected, with the increased expression of Nanog in undifferentiated ESCs versus 3 days +RA differentiating ESCs as measured by RT-PCR (**Figure 3.5A, panel b**). Similarly, the differentiation marker GATA6 was included in the list of genes marked by H3K4me3 in 3 days +RA differentiating ESCs and, likewise, the raw data (**Figure 3.5B, panel a**) confirms that the differentiated sample (bottom line) contained many more sequencing tags that map to the GATA6 locus than did the undifferentiated sample (top line). This result also correlates well with RT-PCR expression data for the GATA6 gene in undifferentiated and 3 days +RA differentiating samples (**Figure 3.5B, panel b**). Notably, the representative genes analyzed by ChIP-RT-PCR and ChIP-seq in **Figure 3.4** above (Oct3/4 and Sox17) do not appear on the lists generated by global bioinformatic analysis. This inconsistency is most likely because of the high stringency used in this initial global analysis, which can produce false negatives.

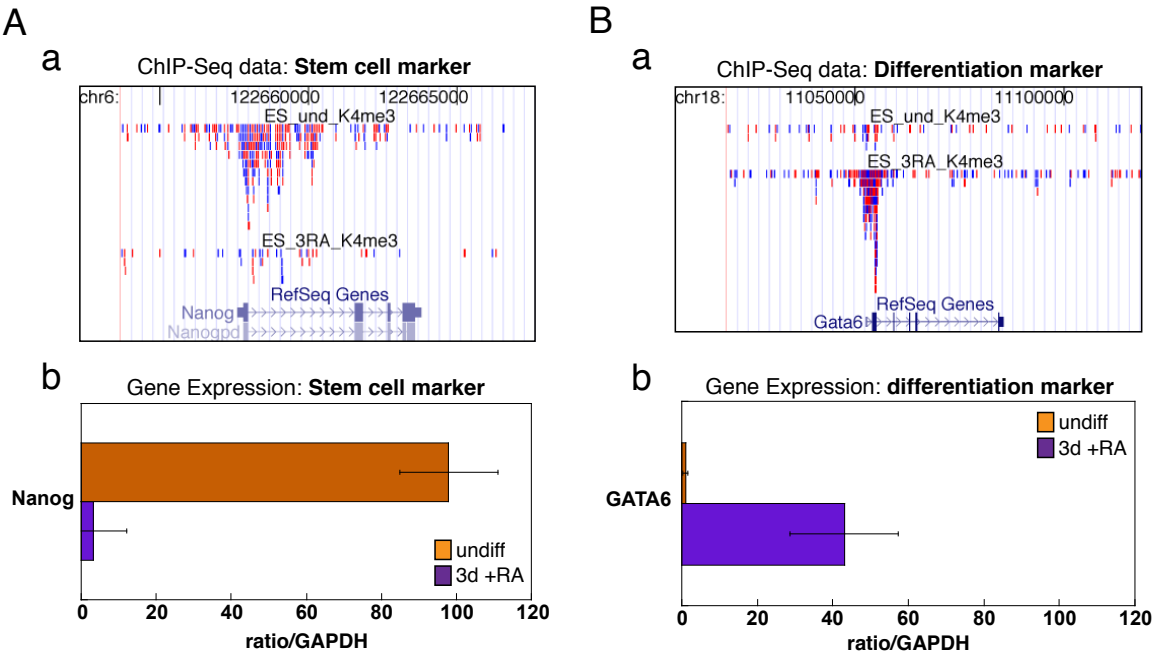


Figure 3.5: H3K4me3 is differentially enriched at functionally different genes in undifferentiated and 3 days +RA differentiating ESCs.

A. H3K4me3 ChIP data for a stem cell marker (Nanog) as assayed by Solexa sequencing (a). Corresponding gene expression (mRNA) data for this marker as assayed by RT-PCR with SYBRgreen (b).

B. H3K4me3 ChIP data for a marker of differentiation (GATA6) as assayed by Solexa sequencing (a). Corresponding gene expression (mRNA) data for this marker as assayed by RT-PCR with SYBRgreen (b).

C. Global analysis of all ChIP-sequence data using Database for Annotation, Visualization and Integrated Discovery (DAVID) functional annotation tools: DNA sequences for each sample (undifferentiated ESCs and 3 days +RA differentiating ESCs) were clustered according to their association with various pathways, assayed for significance (P-value), and ranked accordingly.

D. Global analysis of all ChIP-sequence data using DAVID functional annotation tools: DNA sequences for each sample (undifferentiated ESCs and 3 days +RA differentiating ESCs) were clustered according to their association with various keywords, assayed for significance (P-value), and ranked accordingly.

C

Functional Annotation: Pathways

Undifferentiated ESCs

Term	RT	Genes	Count	%	P-Value	Benjamini
Arachidonic acid metabolism	RT		6	2.0	1.0E-2	8.7E-1
Cell Communication	RT		7	2.4	3.6E-2	9.7E-1
alpha-Linolenic acid metabolism	RT		3	1.0	4.0E-2	9.3E-1
MAPK signaling pathway	RT		10	3.4	4.6E-2	9.0E-1
Linoleic acid metabolism	RT		4	1.4	5.2E-2	8.8E-1

3 days +RA ESCs

Term	RT	Genes	Count	%	P-Value	Benjamini
Neuroactive ligand-receptor interaction	RT		25	4.5	1.7E-6	3.2E-4
Cell Communication	RT		13	2.3	1.0E-3	9.4E-2
Focal adhesion	RT		15	2.7	3.5E-3	2.1E-1
ECM-receptor interaction	RT		9	1.6	5.6E-3	2.4E-1
Calcium signaling pathway	RT		12	2.1	2.5E-2	6.3E-1
Gap junction	RT		7	1.2	7.2E-2	9.1E-1
Cytokine-cytokine receptor interaction	RT		13	2.3	9.4E-2	9.4E-1
Long-term depression	RT		6	1.1	9.6E-2	9.2E-1

D

Functional Annotation: Keywords

Undifferentiated ESCs

Term	RT	Genes	Count	%	P-Value	Benjamini
glycoprotein	RT		70	23.6	2.9E-5	2.5E-2
Cleavage on pair of basic residues	RT		13	4.4	7.2E-5	3.1E-2
signal	RT		57	19.3	6.4E-4	1.7E-1
lipid degradation	RT		5	1.7	1.3E-2	9.4E-1
Secreted	RT		28	9.5	1.4E-2	9.1E-1
structural protein	RT		6	2.0	2.0E-2	9.5E-1
calcium	RT		17	5.7	2.6E-2	9.6E-1
growth factor	RT		6	2.0	3.1E-2	9.7E-1

3 days +RA ESCs

Term	RT	Genes	Count	%	P-Value	Benjamini
Homeobox	RT		46	8.2	6.6E-23	5.8E-20
Developmental protein	RT		60	10.7	4.4E-15	1.9E-12
dna-binding	RT		90	16.1	1.7E-14	4.8E-12
glycoprotein	RT		145	25.9	4.7E-11	1.0E-8
cell adhesion	RT		36	6.4	9.0E-11	1.6E-8
Transcription regulation	RT		74	13.2	4.6E-9	6.7E-7
Secreted	RT		67	12.0	2.2E-7	2.7E-5
Transcription	RT		70	12.5	2.4E-7	2.6E-5

Figure 3.5 cont'd: H3K4me3 is differentially enriched at functionally different genes in undifferentiated and 3 days +RA differentiating ESCs.

Having verified the validity of the ChIP-seq data, I then used the lists to ask whether H3K4me3 preferentially marks genes involved in specific pathways or functions in undifferentiated and/or differentiating ESCs. Using the Database for Annotation, Visualization and Integrated Discovery (DAVID) tools available online at http://david.abcc.ncifcrf.gov/, I was able to cluster each gene list into functionally relevant categories for comparison (**Figure 3.5C and D**). First, I analyzed the gene lists by their association with Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways (**Figure 3.5C**). Interestingly, 25 genes of the 3 days +RA list grouped significantly (P-value = 1.7E-6) in the "Neuroactive ligand-receptor interaction" pathway (**Figure 3.5C, bottom**). This association agrees with other data that suggests these cells are progressing toward a neuronal lineage. Genes from this list also clustered into pathways related to cell motility, which is consistent with the changes these cells undergo during differentiation. The genes of the undifferentiated ESCs (**Figure 3.5C, top**) formed fewer clusters with less significance (P-values in the range of E-2), perhaps because there are fewer genes in the undifferentiated list. The undifferentiated genes clustered in pathways associated with metabolism, but none that indicate development or differentiation as with the 3 days +RA genes.

Next, I analyzed the gene lists by their association with Swiss-Prot (SP) and Protein Information Resource (PIR) database keywords (**Figure 3.5D**). Again, the genes of the undifferentiated ESCs (**Figure 3.5D, top**) formed fewer clusters with less significance than those of the 3 days +RA gene list, and these clusters are all generally related to cell growth and proliferation. Genes in the 3 days +RA differentiated sample list, however, formed several highly significant groupings (**Figure 3.5D, bottom**), and the most significant amongst them are related to differentiation and development. The most significant clustering of genes from this list are those categorized as Homeobox, a finding that also results from the

clustering of these genes according to protein domain (data not shown). Since Homeobox, or Hox, genes are well-characterized as regulators of development and have been shown to be modified by H3K4me3 previously (Milne et al., 2002), this finding further validates both this ESC model of differentiation and our ChIP-seq method.

Chapter 3 Discussion

Both H3.3 incorporation and H3K4me3 are linked to transcriptional activation (Ahmad and Henikoff, 2002; Ng et al., 2003; Santos-Rosa et al., 2002; Schneider et al., 2004; Sutcliffe et al., 2009). Since the transcriptional profile changes dramatically during the differentiation of ESCs, many groups have logically asked whether these two aspects of chromatin change as well and, if so, whether these changes are correlative or causal. In collaboration with other Allis laboratory members and Assistant Professor Deyou Zheng at Albert Einstein College of Medicine, I showed that both the covalent modification profiles and the specific isoforms of histone H3 change during mouse ESC differentiation and that these changes correlate with differences in the transcriptional profile of these cells. Our findings both agree with and add to other studies of mouse ESC differentiation (Mikkelsen et al., 2007), and therefore help to validate this method of ESC differentiation (Smith, 1991) as a model in which to study changes in histone biology during early development.

The association of histone H3 variant H3.3 with transcriptionally active loci was originally established in *Drosophila*, a genetically tractable system (Ahmad and Henikoff, 2002). Studying the potential differences between the three mammalian H3 variants has been more challenging, both because of the greater number of copies of H3 genes in the mammalian genome and because of the lack of specific H3 variant antibodies. Some studies have used epitope tagging to differentially label each H3 variant and investigate their potential differences in chromatin incorporation and function in mammalian cells (Chow et al., 2005; Daury et al., 2006; Jin and Felsenfeld, 2006; Torres-Padilla et al., 2006); however, this approach does not adequately address the possibility that differences in H3 variant expression, as dictated by their endogenous promotors, may influence their differences in chromatin incorporation and function. Although it has been

shown in *Tetrahymena* that expression of H3.3 outside of S-phase is sufficient for its replication-independent incorporation (Yu and Gorovsky, 1997) and that epitope tagged H3.3 immunoprecipitates a unique chaperone (HIRA) in its complex from that of epitope tagged H3.1 (Tagami et al., 2004), it is still unclear how these chaperone proteins mediate the differential incorporation of the H3 variants. Moreover, the histone H3.3 gene structure is significantly different from that of the canonical replication dependent H3 genes: whereas H3.1 and H3.2 genes are intronless and are transcribed into non-polyadenylated mRNAs with short 3' and 5' untranslated regions (UTRs), H3.3 genes have introns and are transcribed into polyadenylated mRNA with lengthy UTRs and an atypical codon usage pattern (Wells et al., 1987). It may therefore be critical to study the expression, incorporation and function of epitope-tagged H3.3 protein under conditions in which it is expressed from its endogenous promoter. Such studies are currently being conducted by my fellow graduate student Aaron Goldberg.

As described above, we chose to use a proteomic approach to study the composition of histone H3 in chromatin in various mammalian cell lines. Chromatography and gel electrophoresis enabled us to separate the three H3 variants and then confirm their identity by mass spectrometry. Although this method does not provide information about the differential association of each H3 variant with specific genomic loci, it did allow us to gain some valuable information about the differences in H3 variant incorporation between various mammalian cell lines and stages of ESC differentiation. ESCs and neuroblastoma (N2a) cells were observed to have a higher relative amount of H3.3 than other cell lines. Moreover, the relative amount of H3.3 generally increased as ESCs differentiated. Interestingly, each of these observations is consistent with a particular theory of H3.3 function or deposition, yet together they are potentially contradictory. The first observation, that the pluripotent (ESCs) or multipotent (N2a) cells have relatively

more H3.3 variant protein than the differentiated cell lines tested, fits a model in which H3.3 is associated with genes that are “poised” for activation: stem-like cells are presumed to have many such loci so that they are ready to progress toward any given lineage upon induction. However, it is also true that these cells are rapidly proliferating and therefore spend a higher percentage of time in S-phase, when the canonical H3 variants (H3.1 and H3.2) should equally be available for deposition. To this end, our second observation, that the relative amount of H3.3 generally increases as ESCs differentiate and slow their proliferation, fits with a model in which amitotic cells accumulate H3.3 (relative to H3.1 and H3.2) since it can be deposited it into chromatin outside of S-phase. However, as noted above, other even more differentiated cell lines (e.g. Raji or U2OS) have relatively less replication-independent H3.3 than do the ESCs. Although there are many trivial explanations for these apparent contradictions (e.g. differences between primary and transformed cell lines), it is also possible that H3.3 has functions in addition to its association with transcription. Moreover, it may be that H3.3 has a unique and specialized function(s) in ESC differentiation and/or early mammalian development that has yet to be identified. This question is also being addressed by my colleague Aaron Goldberg.

Trimethylation of H3K4 is another important marker of active transcription, and one that has potentially important implications in development since it is catalyzed by trithorax proteins, which are required for the proper expression of key developmental regulators such as Hox genes (Ringrose and Paro, 2004). Several groups have examined the H3K4me3 profile, and that of the PcG mediated H3K27me3 modification, in ESC models to see whether it contributes to the maintenance and/or loss of pluripotency (Azuara et al., 2006; Bernstein et al., 2006; Mikkelsen et al., 2007). As expected, and in agreement with my own results described above, these studies find that H3K4me3 marks the transcription

start site of genes required for pluripotency in undifferentiated ESCs and of genes that regulate specific lineages in differentiating ESCs. However, Bernstein et al and Azuara et al additionally found that H3K4me3 was also present on genes that expressed at very low levels. These genes were dually marked by both H3K4me3 and H3K27me3, suggesting that the repressive H3K27me3 modification is dominant over the activating H3K4me3 mark. Bernstein et al termed these loci "bivalent domains", in which large regions of H3K27me3 marked chromatin contain smaller regions of H3K4me3. Azuara et al further demonstrate the dominance of H3K27me3 over H3K4me3 by showing that the transcription at these "bivalent" genes increases in ESCs lacking H3K27me3 (Eed null ESCs, which cannot methylate H3K27 since the lack of Eed disrupts the HMTase activity-containing PRC2 complex). Bernstein et al hypothesize that this bivalent status enables ESCs to keep key developmental regulators in a "poised" state so that they are ready for activation upon differentiation.

Importantly, these studies not only reveal that ESCs have a unique modification profile but also that H3K4me3 may have a unique role in pluripotent cells. Nevertheless, since they were done using ChIP (**see Figure 3.3**) and were therefore constrained by the antibodies chosen and/or available, they may have missed other modifications or patterns of modifications (on histone H3 or another histone) that are potentially unique and important to ESC chromatin. Although here I used the fact that H3K4me3 is a well-established mark to validate my method of ESC differentiation, in the future it would be interesting to use a non-biased approach to assess the global modification patterns on all histones (and histone variants) in ESCs as they differentiate towards different lineages. Such an investigation could potentially reveal other "bivalent" marks, perhaps on or between other histone tails. Two complementary approaches could be used to explore this possibility. First, one could use a proteomic strategy such as that

used by Taverna et al (Taverna et al., 2007) to show long-distance combinatorial methylation and acetylation modifications on the H3 tail in *Tetrahymena*. In this approach, chromatographic methods would be used to separate differently modified histone tails after appropriate digestion with an enzyme that leaves the histone tail intact (e.g. GluC to digest H3, as described in Chapter 4). Mass spectrometry can then be employed to determine the precise combination of modifications on the proteins that elute from the stationary phase at a given time. By comparing the modification profiles of undifferentiated ESCs to that of differentiated ESCs (and repeating with multiple ESC lines and methods of differentiation) it would be possible to identify modification patterns that correlate with the maintenance and/or loss of pluripotency. Second, one could use a ChIP approach with antibodies selected based on either theoretical hypotheses or the data generated using the proteomic approach. The second strategy would then allow one to compare the genomic targets of any novel "stem cell" marks to those mapped by Bernstein et al using H3K4me3 and H3K27me3 ChIP. It may also be important to use functional annotation tools, such as the Database for Annotation, Visualization and Integrated Discovery (DAVID) tools described above, to compare the gene targets of these modifications since they may regulate common pathways even if they are not enriched at identical genomic loci.

Epigenetic regulation is necessary for both the self-renewal and proper differentiation of ESCs; this regulation is accomplished by altering structural and biochemical aspects of chromatin. Here I show that my adaptation of a method of ESC differentiation (Smith, 1991) produces changes in histone H3. These results both support and expand upon the findings of others (Lennox and Cohen, 1988; Mikkelsen et al., 2007; Pina and Suau, 1987), and help to validate this method as a model of ESC differentiation. In the next chapter, I will describe another change that occurs to H3 during ECS differentiation.

Chapter 4: Study of histone H3 proteolysis in ESCs

Summary

When screening samples from a time course of RA-induced differentiating ESCs for changes in histone modifications, as described in Chapters 2 and 3, I observed that histone H3 was proteolytically cleaved at its N-terminus during differentiation. In this chapter, I will describe how I mapped the sites of H3 cleavage (in collaboration with the Hunt laboratory at The University of Virginia) and identified Cathepsin L as at least one protease responsible for proteolytically processing the N-terminal H3 tail. Although cathepsin, and Cathepsin L in particular, are typically thought to reside in the lysosome, our study supports those of others (Goulet et al., 2004; Hiwasa and Sakiyama, 1996) by showing cathepsin activity in the nucleus. Such regulated, limited proteolysis is a mechanism by which the cell can irreversibly remove modifications present on the H3 tail as well as trigger downstream functional effects via chromatin remodeling and/or affecting the affinity of non-histone chromatin proteins (e.g. Polycomb/CBX, discussed in Chapter 2). In addition, I present some data that suggests H3 cleavage may be regulated by covalent modifications present on the histone tail itself. These studies underscore the intriguing possibility that histone proteolysis, brought about by Cathepsin L and potentially other family members, plays a role in development and differentiation that was not previously recognized (Duncan et al., 2008).

Results

A Faster Migrating H3 Species Is Detected in Differentiating Mouse ESCs

To survey changes in histone proteins and their modifications during mouse ESC differentiation, I used immunoblotting to probe whole-cell extracts (WCEs) with various histone antibodies. When probing with specific histone H3 antibodies (e.g. the H3 general C-terminal, H3K27me2, and H3K27me1 antibodies), I observed reproducibly a faster migrating band in samples taken at time points corresponding to days two and three post-induction with retinoic acid (RA). Notably, this band(s) was observed using an H3 general antibody generated against the C-terminus of histone H3 (**Figure 4.1A and 4.1B, left**), but not with an H3 general antibody generated against the first six N-terminal amino acids (**Figure 4.1B, right**). The faster migrating H3 species was also observed when probing immunoblots with an H3-K27me2 antibody (**Figure 4.1A**); in contrast, it was not recognized when replicate immunoblots were probed with the H3-K4me3 antibody (**Figure 4.1A**). Taken together, the results of these experiments suggested that an extreme amino-terminal fragment of H3 was missing in the faster-migrating H3 sub-band.

I then asked if the H3 sub-band was chromatin associated. Micrococcal-digested chromatin was prepared by standard methods (Wysocka et al., 2001) from both undifferentiated ESCs and ESCs undergoing differentiation with RA, and soluble mononucleosomes were probed with the H3 general antibody. As shown in **Figure 4.1C**, the faster migrating H3 band was seen in the undigested chromatin pellet as well as in the solubilized mononucleosomes derived from differentiating cells, but not the chromatin isolated from undifferentiated ESCs (**Figure 4.1C**).

I next asked if the appearance of the faster migrating H3 species was dependent on the methods employed to trigger ESC differentiation. To address this question, ESCs were differentiated using three different methods: monolayer

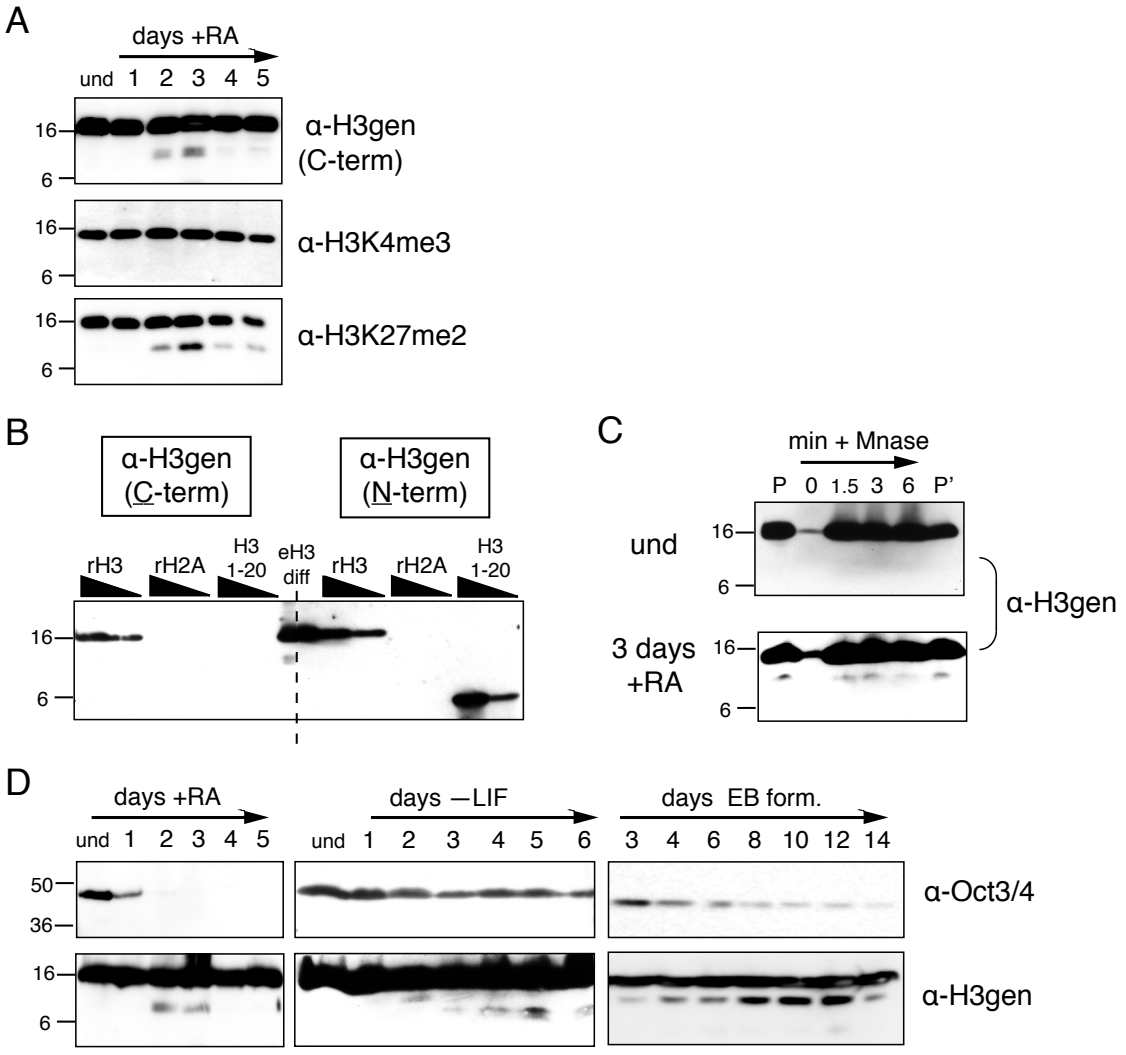


Figure 4.1: A distinct histone H3 species is detected in chromatin during ESC differentiation.

A. Undifferentiated (und) ESCs were differentiated with RA in a monolayer, harvested for WCEs at the time points indicated, and analyzed by immunoblotting with the antibodies indicated to the right of each panel; H3gen will refer to the H3 general C-terminal antibody, unless otherwise indicated. Molecular weights (in kD) are indicated to the left.

B. Recombinant(r) histone H3, rH2A, H3 1-20 peptide and RP-HPLC purified endogenous H3 from differentiated ESCs (eH3 diff) were analyzed by immunoblotting; although the H3gen (C-terminal) antibody recognized the faster migrating H3 sub-band, an H3 N-terminal antibody does not.

C. Chromatin isolated from either undifferentiated ESCs (top panel) or those differentiated with RA for 3 days was subsequently digested with micrococcal nuclease for the indicated times; the solubilized chromatin pellet input (P), Mnase digested chromatin, and solubilized post-Mnase pellet (P') was then analyzed by immunoblotting with the H3-gen antibody.

D. ESCs were differentiated using three basic methods: monolayer differentiation with RA (left), monolayer differentiation with LIF withdrawal (middle), and embryoid body (EB) formation by cell aggregation (right); WCEs were analyzed as in **A** for both a marker of pluripotency (Oct 3/4, top panels) and the histone H3 sub-band.

differentiation with RA, monolayer differentiation by withdrawal of leukemic inhibitory factor (LIF), and embryoid body formation (EB formation) by cell aggregation. As shown in **Figure 4.1D** (upper panels), the time course for expression of pluripotency marker Oct3/4 differs for each of the induction methods employed, suggesting differences in the timing and progression of differentiation. Timing of the appearance of the faster migrating H3 band was also dependent on the method used to induce ESC differentiation (**Figure 4.1D**, lower panels), suggesting that this event is also dependent on the progress of differentiation. The faster migrating H3 was observed at days two and three post RA induction during monolayer differentiation (**Figure 4.1A, upper panel, and Figure 4.1D, lower left panel**), but delayed until day five following withdrawal of LIF (**Figure 4.1D, lower middle panel**), the latter correlating with the similarly delayed decrease of Oct3/4. During EB formation, the faster migrating H3 species appeared early and then peaked between days eight and twelve (**Figure 4.1D, lower right panel**), suggesting a slower and more complex differentiation progression. I chose to use monolayer differentiation with RA for subsequent experiments in order to minimize differences in the timing and heterogeneity of differentiation.

Histone H3 Proteolysis Does Not Correlate With Markers of Apoptosis

To address the question of whether this proteolysis is the result of apoptosing cells rather than differentiating cells, I probed both +RA differentiating cells and cells induced to die with markers of apoptosis and DNA damage. As shown in **Figure 4.2A**, ESCs treated with RA did not show strong evidence of caspase 3 activation. However, since ESCs treated with the apoptosis inducers etoposide (VP16) and staurosporine did not show caspase 3 activation either, I also treated HL60 cells with etoposide to verify that this drug could indeed induce apoptosis.

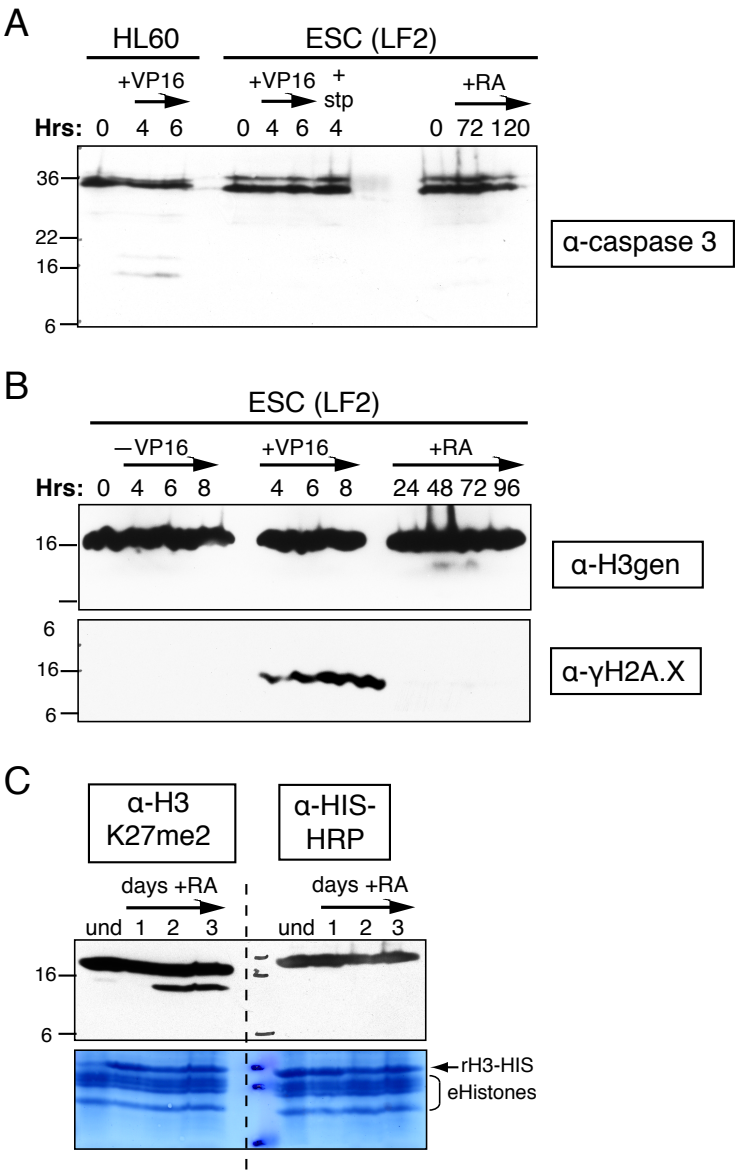


Figure 4.2: Histone H3 cleavage does not correlate with markers of apoptosis or DNA damage.

A. Human promyelocytic leukemia (HL60) cells and ESCs were treated with either etoposide (VP16), staurosporine (stp), or retinoic acid (RA) for the indicated time. Although HL60 cells showed caspase 3 activation upon VP16 treatment, ESCs did not show significant caspase 3 activation under any of the conditions tested.

B. ESCs treated with VP16 do show upregulation of the DNA damage marker γ-H2A.X, but do not show H3 cleavage. ESCs treated with RA do show H3 cleavage, as described above, but do not show upregulation of γ-H2A.X.

C. To test whether uncontrolled proteolysis occurred upon cell lysis, unmodified rH3-HIS was added to SDS-Laemmli sample buffer prior to adding it to cell pellets and solubilization to generate WCEs. The unmodified rH3-HIS showed no evidence of cleavage when added to the lysing cells (right); however, endogenous H3 cleavage was detected in the lysates of 2 and 3 day differentiated ESC, as expected, which was shown by immunoblotting with an H3K27me2 antibody (left).

The etoposide did cause apoptosis of the HL60 cells, as shown by the appearance of caspase 3 activation (indicated by its cleavage to the smaller products migrating at ~16kD). Interesting, ESCs treated with etoposide, but not RA, did show signs of DNA damage, as indicated by the induction of γH2A.X signal (**Figure 4.2B**). These results are consistent with reports in the literature showing that ESCs are more resistant to apoptosis than other cell types, and that they have greater activation of mechanisms to repair DNA damage.

Recombinant Histone H3 Is Not Cleaved "Ex Vivo"

To test whether proteolysis occurred upon cell lysis, I added unmodified rH3-HIS to SDS-Laemmli sample buffer prior to adding it to cell pellets and then proceeded as usual. As shown in **Figure 4.2C**, the unmodified rH3-HIS showed no evidence of cleavage when added to the lysing cells (**right**); however, the lysates did show evidence of endogenous H3 cleavage, as shown by immunoblotting with an α-H3-K27me2 antibody (**Figure 4.2C, left**).

Histone H3, Marked by Both "Active" and "Silent" Modifications, Is Proteolytically Cleaved in The N-terminal Tail During ESC Differentiation

To determine the nature of the faster migrating H3 sub-species, I extracted histones from differentiating ESC nuclei three days post RA induction and separated the four core histone proteins by reverse phase high pressure liquid chromatography (RP-HPLC, C8 column, as in **Figure 3.1B**). Fractions containing the H3 sub-band were pooled and further resolved by RP-HPLC (C18 column, **Figure 4.3A**), and the resulting fractions were screened by immunoblot as shown in **Figure 4.3B** (left panel). I then subjected C18-RP-HPLC fractions containing the H3 sub-band to two separate sequencing methods. First, I pooled equal amounts of fractions 52-54, separated the protein by SDS-PAGE, and transferred it to immunoblot; the

Figure 4.3: Histone H3 is N-terminally cleaved during ESC differentiation.

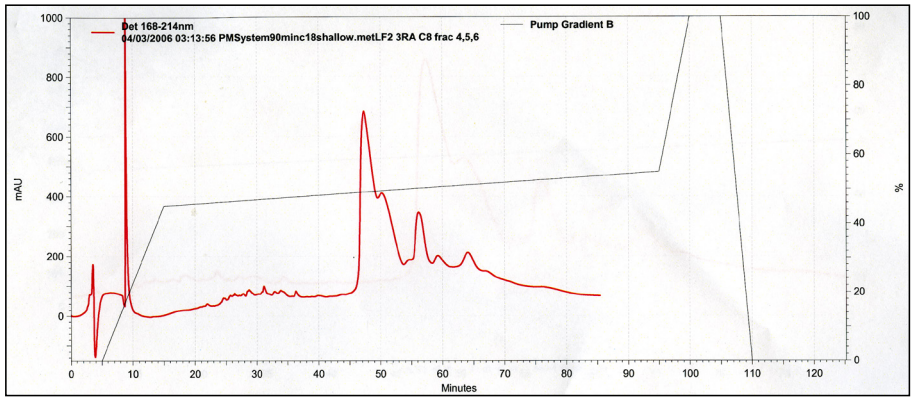
A. Total histones were acid-extracted from 3 days +RA differentiating nuclei and purified using RP-HPLC with a C8 column (not shown); fractions containing the histone H3 sub-band were pooled and re-fractionated by RP-HPLC with a C18 column (chromatogram).

B. Fractions eluted from the C18 column (as shown in **A**) were then screened by immunoblotting with the H3-gen antibody (left). Equal amounts of fractions 52-55 were pooled, separated by SDS-PAGE, transferred to PVDF membrane and stained with Ponceau Red (middle). Both bands of the sub-band doublet (asteriks) were excised from the membrane and subjected to Edman degradation, which provided evidence that the faster migrating H3 had been N-terminally cleaved between residues A21 and T22 and between K27 and S28 (right). Residues not clearly identified in the generated sequence are denoted "X."

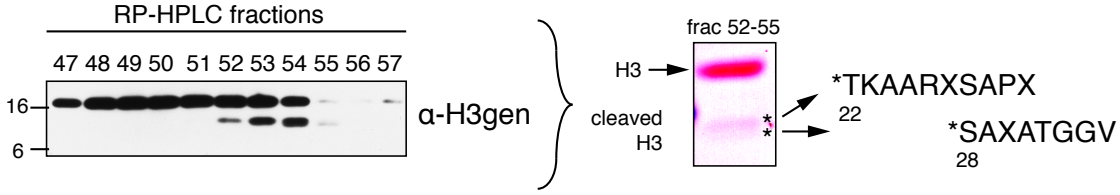
C. Sample in fraction 54 from the RP-HPLC enrichment (screened in **B**) was digested with GluC, which generates intact N-terminal peptides that terminate with E50. Peptides generated were then analyzed on a linear ion trap–Fourier transform mass spectrometer. Six highly modified, truncated fragments of the GluC-generated 1-50 peptide were observed (right panel; detected post-translational modifications are summarized in **Figure 4.4**). Summed ion currents for all charge states and modified forms of the above sequences were employed to estimate the relative abundances of the six peptides beginning with the indicated residues as follows: T22 and A24 > than K23 and K27 > A25 and S28 (see right hand list). The peptide beginning with R26 was not detected. Note that the two sequences detected by Edman degradation were also detected by MS (see asterisks). All six of the truncated 1-50 peptides contain Ala at position 31 (underlined) and are thus derived from the H3 isoform H3.2. Three highly modified forms of the complementary N-terminal fragments, A1-A21, A1-K23, and A1-R26 (left hand list) are also present in the same HPLC fraction.

D. The sequence of the mammalian histone H3 tail and the cleavage sites mapped in **A**-**C**; the bold solid line indicates the "primary" cleavage site mapped by both Edman degradation and MS (H3. cs1); additional significant cleavage sites are marked with regular solid lines; less abundant sites are marked by dashed lines. Lysines found by MS to be highly acetylated (ac) or methylated (me) are marked by a triangle or circle.

A



B



C

MS analysis of endogenous histone H3 cleavage

N-terminal peptides		C-terminal peptides	
1-21	ARTKQTARKSTGGKAPRKQLA	*TKAARKSAPATGGV…	22-50
1-22	n.d.	KAARKSAPATGGV…	23-50
1-23	ARTKQTARKSTGGKAPRKQLATK	AARKSAPATGGV…	24-50
1-24	n.d.	ARKSAPATGGV…	25-50
1-25	n.d.	n.d.	26-50
1-26	ARTKQTARKSTGGKAPRKQLATKAAR	KSAPATGGV…	27-50
1-27	n.d.	*SAPATGGV…	28-50

data in C produced by Tara Muratore-Schroeder and used with permission

D

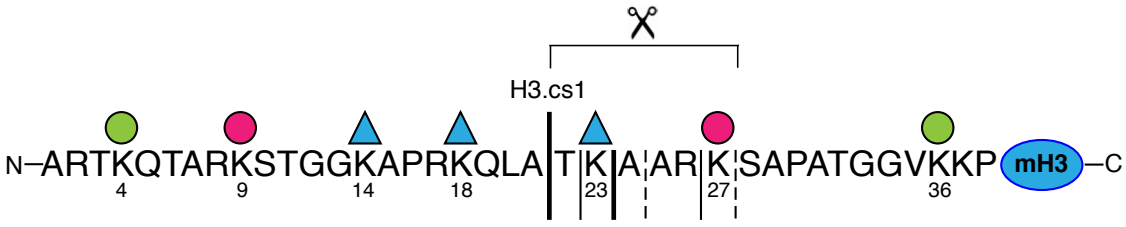


Figure 4.3: Histone H3 is N-terminally cleaved during ESC differentiation

material in each of in the two sub-bands labeled by asteriks were then sent to Richard Cook at Baylor University and subjected to Edman degradation (**Figure 4.3B, right**). Although multiple amino acids were released during each cycle for both samples, the observed data strongly suggested that the top and bottom sub-bands contained sequences derived from cleavage following residues A21 and K27, respectively, in the N-terminal tail of H3.

Second, to define in greater detail the peptide sequences and post-translational modifications present in the faster migrating H3 species, additional sample from fraction 54 (**Figure 4.3B**) was sent to the Hunt Laboratory at The University of Virginia where they digested it with GluC to produce N-terminal H3-fragments ending in E50; the resulting peptides were then analyzed by MS. Spectra recorded detected six, highly-modified, truncated, N-terminal, histone H3-peptides beginning at residues T22, K23, A24, A25, K27 and S28 (**Figure 4.3C, right panel**). Relative abundances of the six peptides suggest that the preferred cleavage sites are C-terminal to A21 and K23. These results suggest that the primary H3 cleavage site is between amino acids 21 and 22 of the amino terminus (**Figure 4.3D, H3.cs1**) and that the final cleavage site is between amino acids 27 and 28. Notably, the list of cleavage sites detected by MS contains the two sites detected by Edman degradation (asterisks).

In addition, new mass spectrometric methods using a combination of electron transfer dissociation/proton transfer charge reduction and accurate mass measurements were employed by the Hunt laboratory to characterize the modification patterns on the proteolyically cleaved H3 (Coon et al., 2005; Taverna et al., 2007). Interestingly, the data revealed that the cleaved H3 sub-species has a distinct covalent modification profile, suggesting that the H3 sub-band may be preferentially marked, before or after proteolytic processing, with a specific epigenetic signature (**Figure 4.3D**). Specifically, marks of both “active” (e.g. H3K23Ac and

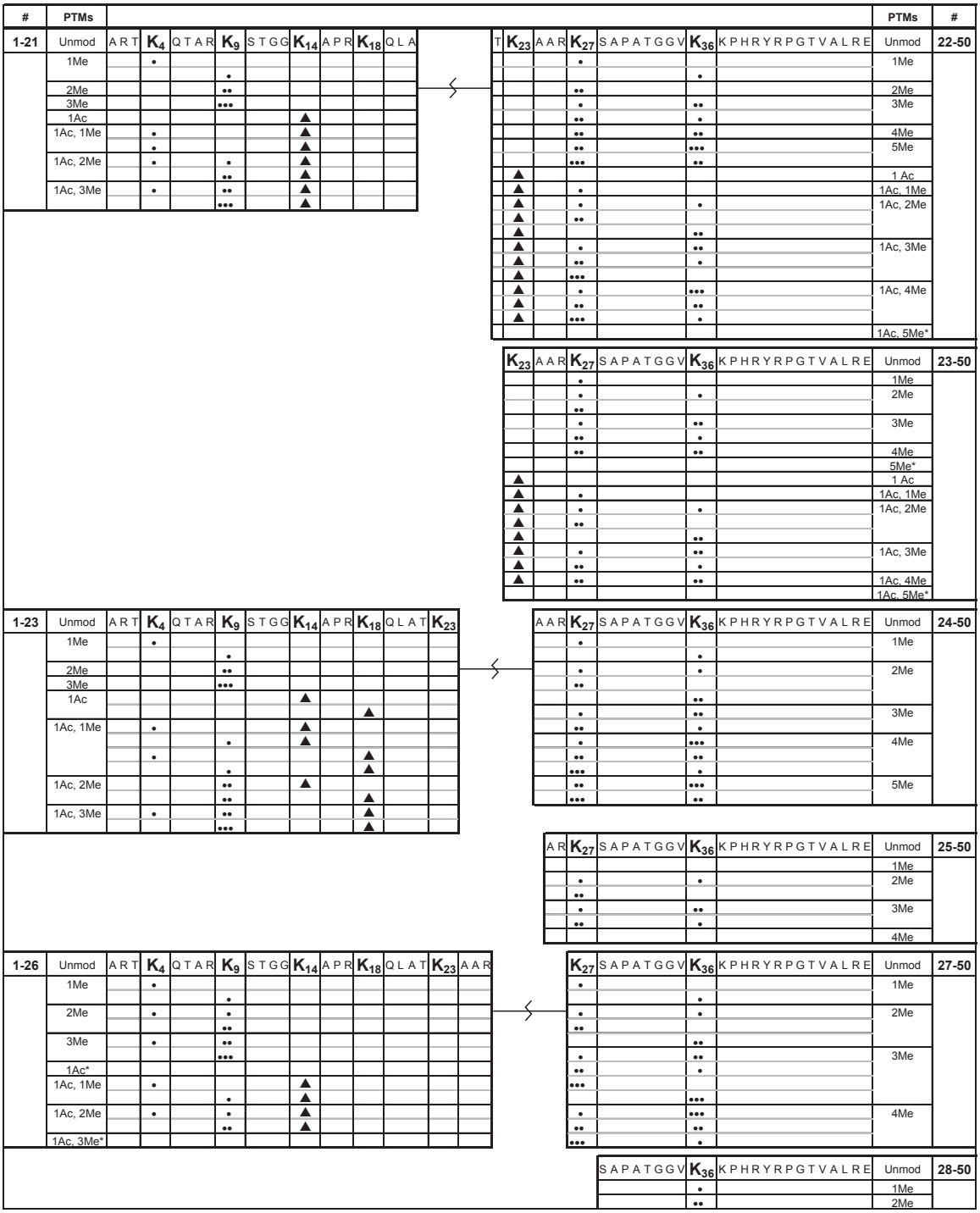
H3K36me) and “silent” transcription (e.g. H3K27me) were reproducibly detected on a single GluC-digested peptide derived from the proteolytically-processed H3 fragment. Moreover, all six of the truncated peptides contained Ala at position 31, revealing that the cleaved H3 peptide is H3 isoform H3.2, not H3.3. Although non-cleaved H3.3 peptide was detected in the same RP-HPLC fraction as the cleaved H3 species, cleaved H3.3 was not detected. Since H3.2 and H3.1 elute in two separate peaks (**Figure 4.3B**), we cannot definitively conclude that H3.2 is preferentially cleaved over H3.1, however significantly less H3 cleavage was detected in peak 2 (H3.1) than in peak 1 (H3.2+H3.3) by immunoblotting. Although it is unclear how this particular pattern of modifications and H3 isoform affect the mechanism of proteolysis, these data suggest the intriguing possibility that that regulation via post-translational modification of the substrate and/or isoform preference may regulate the proteolytic processing (**see Figure 4.12**).

Also detected by MS in fraction 54 (**Figure 4.3B**) were highly modified forms for three of the most abundant, complementary, N-terminal fragments generated by proteolytic cleavage of H3 (**Figure 4.3C, left panel**). The detection of these intact, complementary cleavage products indicates that the cleavage sites mapped above are the result of endopeptidase activity alone. Post-translational modifications detected on these three N-terminal peptides again include marks of both “active” (e.g. H3K14Ac) and “silent” transcription (e.g. H3K9me) as seen on the C-terminal fragments (**Figure 4.4**). These findings support the conclusion that a small fraction of total histone H3.2 undergoes highly specific endoproteolytic cleavage during ESC differentiation that may be regulated by unique patterns of covalent modifications.

The detection of the N-terminal cleavage product was unexpected, given that the ESC histones were acid-extracted and purified over two columns. However, the abundance observed by the Hunt Laboratory for the N-terminal fragments were

Figure 4.4: Summary of the post-translational modifications detected on proteolytically cleaved H3 obtained from differentiating ESCs.

Material from fraction 54 (**Figure 4.3B**) was digested with GluC to produce N-terminal, H3 fragments ending in E50. The resulting mixture was then analyzed by nano-flow HPLC interfaced with both a linear ion trap-Fourier transform mass spectrometer and a linear ion trap instrument equipped for electron transfer dissociation. Spectra recorded with the former instrument detected six, highly-modified, truncated, N-terminal, histone H3-peptides beginning at residues T22, K23, A24, A25, K27 and S28 and ending at E50 (**Figure 4.3C**, right panels). Post-translational modifications detected by recording electron transfer dissociation (ETD) mass spectra on different isoforms of these six peptides are shown above in the right panel. Sixty-seven different forms of the above six peptides were characterized. Ten contain marks associated with active transcription (H3K23Ac, H3K36me, H3K36me2 or H3K36me3), ten contain marks associated with gene silencing (H3K27me, H3K27me2 or H3K27me3), and 43 contain combinations of above active and repressive marks. Summed ion currents for all charge states and modified forms of the above sequences were employed to estimate the relative abundances of the six peptides beginning with the indicated residues as follows: T22 and A24 > than K23 and K27 > A25 and S28 (see right hand list). The peptide beginning with R26 was not detected. Three of the most abundant, complementary, N-terminal fragments generated by proteolytic cleavage of H3 (A1-A21, A1-K23, and A1-R26) were also detected in HPLC fraction 54 (**Figure 4.3C**, left panels). Post-translational modifications detected on different isoforms of these three peptides are shown above in the left panel. Monomethyl-, dimethyl-, and trimethyl-marks are depicted by one, two, and three solid circles, respectively. Acetyl marks are depicted by triangles. An asterisk indicates that the modified isoform was detected by accurate mass measurement only. All ETD spectra were interpreted manually.



produced by Tara Muratore-Schroeder and used with permission

Figure 4.4: Summary of the post-translational modifications detected on proteolytically cleaved H3 obtained from differentiating ESCs.

down by 10 fold compared to that observed for the complementary C-terminal fragments. Moreover, we processed histones from undifferentiated cells in parallel and did not detect any evidence of cleaved histone H3. I also found that the cleavage sites presented were mapped in three separate samples, arguing against the random, artifactual degredation.

The Lysosomal Cysteine Protease Cathepsin L is Present in Fractions That Are Enriched with H3 Protease Activity

To identify and characterize the putative H3 protease(s), I first established an *in vitro* H3 cleavage assay. Undifferentiated or differentiating ESCs were harvested and their proteins extracted; these lysates were then incubated with C-terminally 6xHIS tagged, full-length recombinant histone H3 (rH3-HIS), and the reaction products were analyzed by immunoblotting with the Qiagen Penta HIS-HRP antibody (**Figure 4.5A**). Although this assay has the potential to detect any N-terminal H3 cleavage, chromatin extracts isolated from differentiating ESCs at the time point when *in vivo* cleavage was observed (~3 days +RA) consistently possessed strong *in vitro* H3 cleavage activity that reproduced the size difference of the endogenously cleaved H3 (**Figure 4.5B**).

To better ensure that the cleaved H3 product observed in our *in vitro* cleavage assay was physiologically relevant, I generated an antibody that recognized the primary site of H3 cleavage, H3.cs1 (**Figure 4.3D**), using a 2x branched peptide (**Figure 4.6A**). ELISA results indicated that this antibody ("H3.cs1") was specific for the primary site of *in vivo* generated H3 cleavage product (A21/T22) and was not sensitive to the acetylation status of H3K23 (**Figure 4.6B**). Importantly, the H3.cs1 antibody fails to react with full length H3 and its signal is highly enriched at the same time point at which I detect an H3 sub-band using the H3-general or H3 K27me2 antibodies during our standard RA-induced ESC differentiation time course (**Figure 4.6C**).

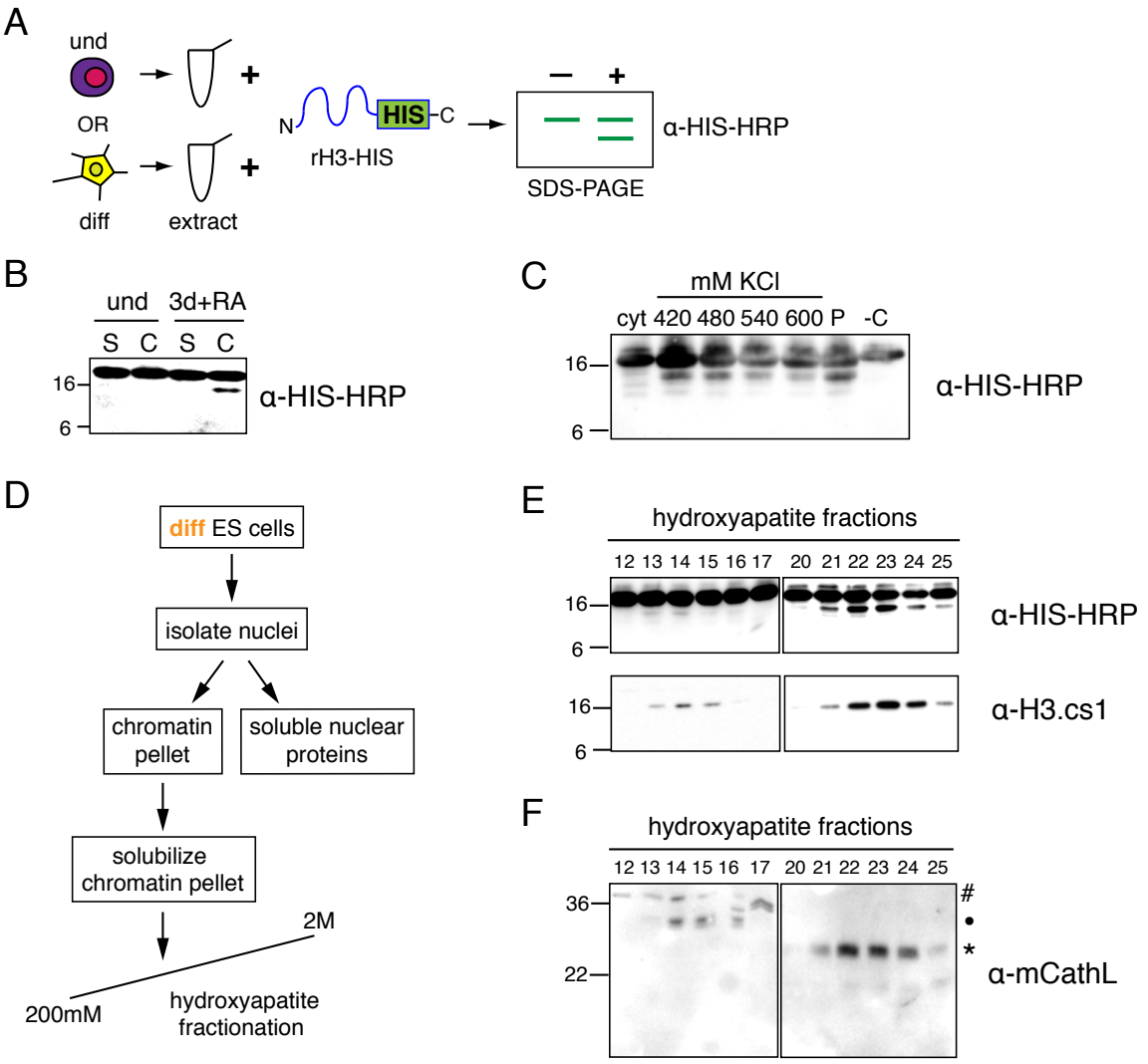


Figure 4.5: The cysteine protease Cathepsin L is detected in fractions enriched for histone H3 cleavage activity.

A. Schematic of *in vitro* H3 cleavage assay (see text for details).

B. An example of the H3 cleavage assay comparing soluble cytosolic + nuclear protein extract (S) and solubilized chromatin extract (C) from undifferentiated and 3 days + RA differentiating ESCs.

C. Nuclear extracts were prepared as described (Dignam et al., 1983) from 3 days +RA differentiated cells; after the initial extraction of nuclear proteins with 420 mM KCl (420), the chromatin pellet (P) was further extracted by sequential 60mM increases in KCl concentration. The remaining chromatin pellet was then solubilized by sonication in buffer A. The cytosolic (cyt), high salt, and chromatin extracts were then assayed for H3 cleavage activity using the H3 cleavage assay described in **A**.

D. Schematic of extract fractionation for protease enrichment.

E. H3 cleavage assay of hydroxyapatite fractions generated by scheme shown in **D**; assay reactions were analyzed by immunoblotting with both HIS-HRP and H3.cs1 antibodies.

F. Hydroxyapatite fractions assayed in **D** were analyzed for the presence of Cathepsin L by immunoblotting with Cathepsin L antibody; # designates proprotein (~37kD), • indicates intermediate processed form (~30kD), and * indicates mature processed form (~25kD).

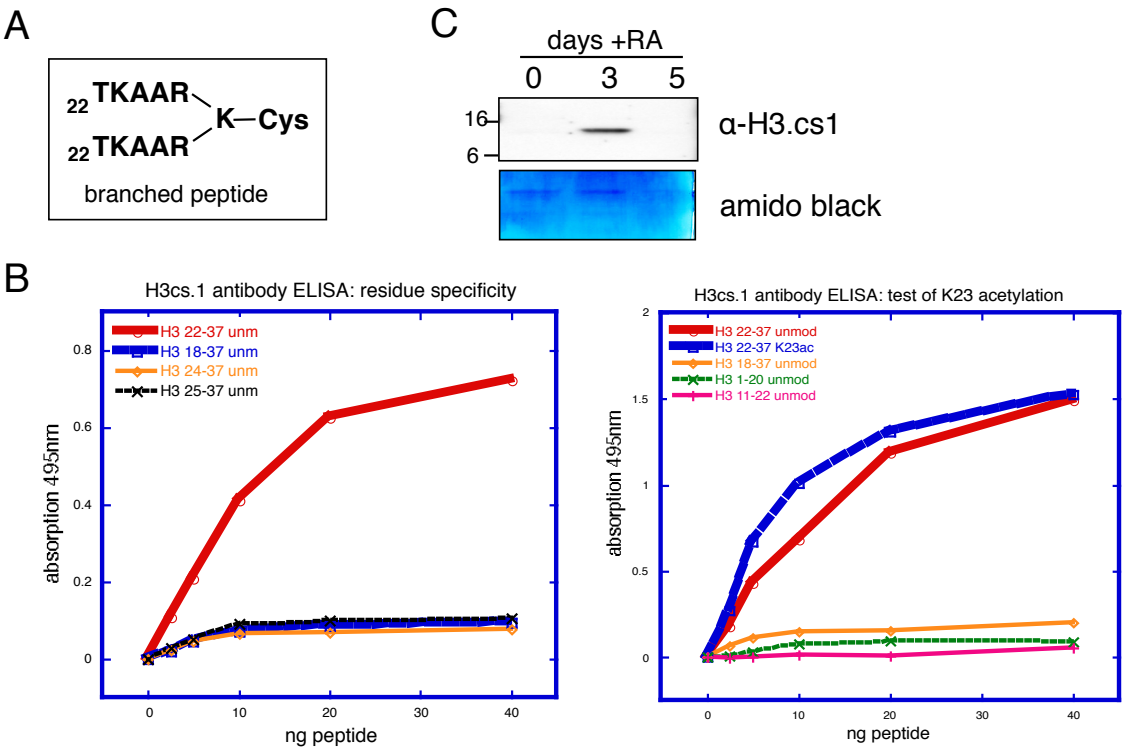


Figure 4.6: Characterization of H3.cs1 antibody.

A. A 2x branched peptide sequence used to generate antibody (see Chapter 6: Materials and Methods for details).

B. Rabbit serum was tested for biological specificity by immunoblotting on WCEs from undifferentiated, 3 days +RA, and 5 days +RA.

C. Rabbit serum was tested for amino acid sequence and modification specificity (H3K23ac) by ELISA.

I then used both the HIS-HRP antibody and the H3.cs1 antibody in the above H3 cleavage assay to follow the biochemical enrichment of H3 protease activity in nuclear extract derived from differentiating ESCs. In the process of generating and testing extracts for protease activity, I noted that the majority of protease activity remained in the chromatin pellet even after multiple high salt extractions (**Figure 4.5C**). For this reason, I sought a method with which I could solubilize, extract and fractionate chromatin directly. I turned to hydroxyapatite resin fractionation since it has been used in the literature for this exact purpose and, in fact, often used to purify histone proteins themselves (Bloom and Anderson, 1978; Simon and Felsenfeld, 1979). I fractionated nuclear extract as described by the schematic in **Figure 4.5D** and detected two peaks of putative H3 cleavage activity by both the

HIS-HRP antibody and the H3.cs1 antibody (**Figure 4.5E**). I then sent two activity-containing fractions (#22 and 23, **Figure 4.5E**) and an adjacent non-activity-containing fraction (#20) to the Hunt Laboratory for MS analysis. Four peptides were detected for the cysteine protease Cathepsin L in both activity-containing fractions 22 and 23, yet none of these peptides were detected in the neighboring non-activity containing fraction 20 (**Figure 4.7**). To validate this identification further, I used a commercially available antibody to screen for the presence of Cathepsin L in the hydroxyapatite fractions (mCathL, **Figure 4.5F**). Importantly, reactivity with mCathL correlates well with the detection of H3 cleavage activity in the fractions aligned above (**compare Figure 4.5E to 4.5F**).

>gil6753558|ref|NP_034114.1| cathepsin L preproprotein
Mus musculus]

MNLLLLLAVLCLGTALATPKFDQTFSAEWHQWKSTHRRLYGTNEEEWRRAIWE
KNMRMIQLHNGEYSNGQHGFSMEMNAFGDMTNEEFRQVVNGYRHQKHKK
GRLFQEPLMLKIPKSVDWREKGCVTPVKNQGQCGSCWAFSASGCLEGQMFLK
TGKLISLSEQNLVDCSHAQGNQGCNGGLMDFAFQYIKENGGLDSEESYPYEAK
DGSCKYRAEFAVANDTGFVDIPQQEKALMKAVATVGPISVAMDASHPSLQFYS
SGIYYEPNCSSKNLDHGVLLVGYGYEGTDSNKNKYWLVKNSWGSEWGMEGYI
KIAKDRDNHCGLATAASYPVVN

Four Peptides Detected
ENGGLDSEESYPYEAK
NSWGSEWGMEGYIK
DRDNHCGLATAASYPVVN
DNHCGLATAASYPVVN

data produced by Tara-Muratore-Schroeder and used with permission

Figure 4.7: Mass spectrometry analysis of hydroxyapatite fractions identifies the cysteine protease Cathepsin L specifically in those fractions exhibiting H3 cleavage activity.
To identify the putative H3 protease, proteins in two of the active fractions (#22 and #23) and one of the adjacent non-active fractions (#20) were digested with trypsin. The resulting peptides were analyzed by nano-flow HPLC interfaced with a linear ion trap-Fourier transform mass spectrometer and several thousand collision activated dissociation (CAD) mass spectra were acquired. By searching theses spectra against a database of murine proteins with the SEQUEST algorithm, we identified more than 1,000 peptides from 80-100 proteins in each fraction. Four low level tryptic peptides (0.5%) detected in each of the two active fractions but not in the inactive fraction, matched to the mature form of the lysosomal cysteine protease, Cathepsin L. All were present at or below the 0.5% abundance level. Sequences for the detected peptides are listed below the full sequence and the corresponding residues are underlined within the full amino acid sequence of Cathepsin L. No other proteases were detected in the above analyses.

Cathepsin L is known to exist in three forms: a proenzyme running at ~37kD, a single chain intermediate at ~30kD, and a double chain mature form at ~25kD and ~5kD (Ishidoh et al., 1998). The pro form must be cleaved to become active, whether by self-cleavage or by another enzyme (Turk et al., 2000); in contrast, both the "intermediate" and "mature" forms have been shown to be active (Mason et al., 1989). The hydroxyapatite fractions of the stronger peak of activity (~21-25) correlate with the detection of the mature form of Cathepsin L by immunoblot (**Figure 4.5F**, marked by asterisk), while the weaker activity fractions (~13-15) correlate closely with the detection of the intermediate form of Cathepsin L (**Figure 4.5F**, marked by dot). This correlation not only supports a causal relationship between the presence of Cathepsin L and the H3 cleavage activity, but also helps to explain the difference in elution and activity between the two peaks of H3 cleavage activity generated by the hydroxyapatite chromatography.

Fractions Enriched in H3 Cleavage Activity Exhibit Cathepsin L-like Activity

To validate further the identification of Cathepsin L as a histone H3 protease, I tested our enriched fractions with different classes of protease inhibitors in our H3 cleavage assay. As shown in **Figure 4.8A**, adding increasing amounts of serine protease inhibitor AEBSF produced a modest reduction in cleavage activity at a high concentration (20mM), but not at the lower concentrations tested (2mM and 10mM). However, when using the cysteine protease inhibitor E64, all cleavage was abolished at even the lowest concentration tested (10μM).

Noting the strong inhibition of H3 cleavage activity by E64, an irreversible inhibitor that binds covalently to its substrate and has been shown to inhibit Cathepsin L (Barrett et al., 1982), I then took advantage of a commercially available E64-bound resin to test whether it could precipitate Cathepsin L from our active hydroxyapatite fractions and, subsequently, remove the soluble H3 cleavage activity. Both E64 bound and control resins were added to fractions 20 (in which

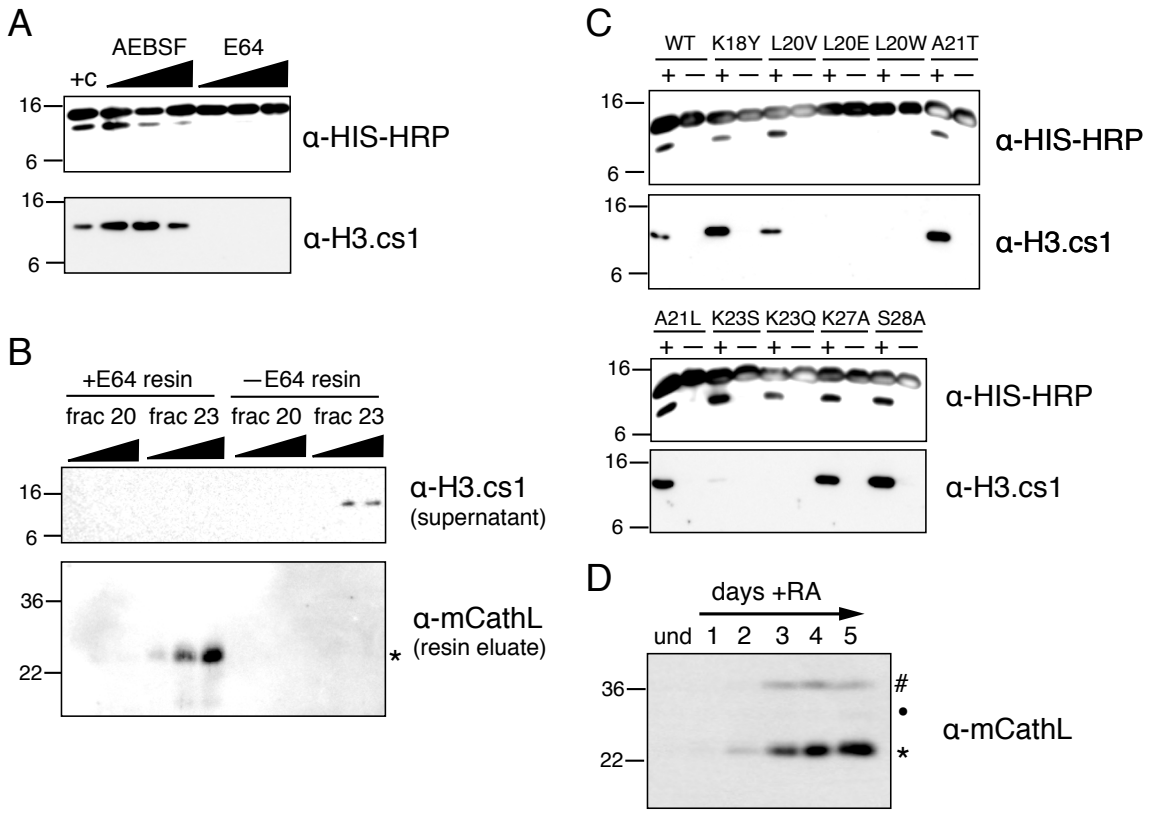


Figure 4.8: Cathepsin L cleaves histone H3 *in vitro* and associates with chromatin *in vivo*.
A. Hydroxyapatite fraction #23 (**Figure 4.5E**) was assayed +/- protease inhibitors in the H3 cleavage assay; cysteine protease inhibitor E64 is a potent inhibitor of the H3 protease activity.
B. Immobilized E64 was incubated with both an active hydroxyapatite fraction (23) and one without enzymatic activity (20); control resin was incubated with each fraction in parallel. Resins were then precipitated from solution, boiled in SDS sample buffer to remove bound proteins, and analyzed by immunoblotting (bottom panel); the supernatant was tested for H3 protease activity (top panel).
C. Hydroxyapatite fraction #23 (**Figure 4.5E**) was assayed with various rH3-HIS point mutants in the H3 cleavage assay; note that mutations in L20 abolished activity as assayed by both the HIS and H3.cs1 antibodies (top two panels) and that mutations in K23 abolished cleavage as assayed by the H3.cs1 antibody (bottom panel).
D. Chromatin from undifferentiated ESCs and ESCs differentiated +RA for the number of days indicated was digested with micrococcal nuclease and analyzed by immunoblotting with Cathepsin L antibody to assay whether or not Cathepsin L protein is associated with chromatin; note that Cathepsin L is associated with chromatin in differentiating ESCs, particularly the mature form (*).

---

little to no H3 cleavage activity detected) and 23 (in which strong H3 cleavage activity was detected), incubated to allow binding and potential inhibition, and then pelleted by centrifugation. The cleared supernatant was subsequently tested for H3 cleavage activity. As expected, E64 resin successfully removed H3 cleavage activity from fraction 23, while control resin did not (**Figure 4.8B**, upper panel).

Importantly, as shown using the cleavage-site specific H3.cs1 antibody, the activity removed by the E64 resin included the specific A21/T22 cleavage site activity. In contrast, the specific H3 cleavage activity remained in the supernatant of fraction 23 incubated with control resin (**Figure 4.8B**, upper panel).

I then eluted any bound proteins from the E64 and control resins by boiling in SDS sample buffer and probed the eluates for the presence of Cathepsin L. As hypothesized, mCathL was detected on the +E64 resin that had been incubated with fraction 23, but was not found on control resin nor on resin incubated with fraction 20 (**Figure 4.8B**, lower panel). As a control, I also analyzed the above eluates by immunoblotting with an antibody to another cathepsin family member, Cathepsin B, but did not detect any immunoreactive species (data not shown). Taken together, the loss of activity from solution paired with the presence of Cathepsin L protein bound to the corresponding +E64 resin strongly suggests a causal relationship between this cysteine protease and the primary H3 protease activity characterized above (cleavage of H3 between A21/T22, H3.cs1).

Cathepsin L is also known to preferentially cleave proteins that contain hydrophobic residues in their P2 position (two residues N-terminal from the cleaved bond, as originally defined (Schechter and Berger, 1967)), specifically leucine and phenylalanine (Rawlings and Barrett; Rawlings et al., 2008). Importantly, and in keeping with this characteristic, the sequence surrounding the primary H3 cleavage site (H3cs.1) mapped from ES cells by both MS and Edman degradation includes a leucine at the P2 position (L20). To verify the significance of this leucine in the H3 sequence and support the identification of Cathepsin L in fraction 23, I mutated this residue (L20V, L20E & L20W) and neighboring residues and tested the mutated recombinant proteins in our H3 cleavage assay (**Figure 4.8C**). Although the more conservative L20V mutant showed little change in H3 cleavage, L20E and L20W mutants completely abolished H3 cleavage as assayed by both the

HIS and H3.cs1 antibodies (**Figure 4.8C**, top two panels). Moreover, the loss of H3 cleavage upon mutation of L20, compared to little or no loss of signal upon mutation of neighboring residues (K18 or A21), suggests that L20 is particularly important to the enzymatic activity in fraction 23. Together, this mutational analysis also supports the identification of Cathepsin L as the enzyme responsible for the histone H3 cleavage characterized in the above *in vitro* assays.

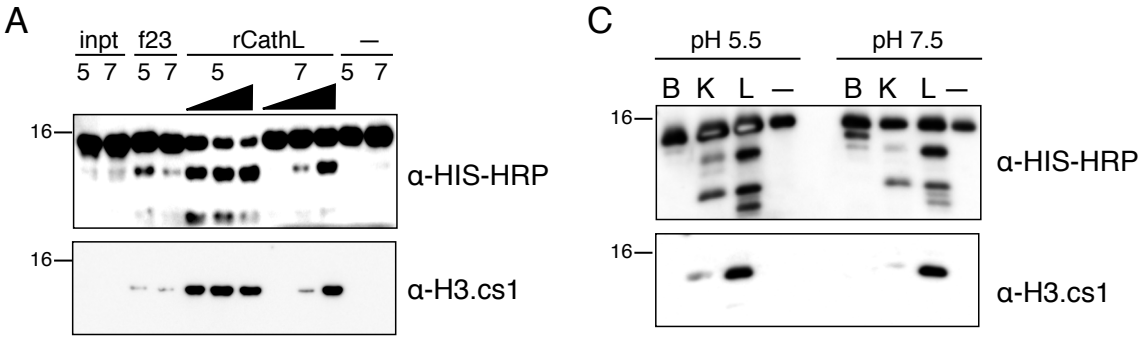
Interestingly, mutating K23 to either S or Q also had a significant effect on the H3 cleavage activity. Although no difference was seen in H3 cleavage by assaying with HIS, these mutations greatly diminished the cleavage site recognized by the H3.cs1 antibody, suggesting a possible role for this residue in regulating the precision of H3 proteolysis by Cathepsin L. It is interesting to note that none of the N-terminal peptides identified by MS were acetylated at K23 while C-terminal peptides acetylated at K23 were consistently detected (**Figure 4.4**). Together, these data suggest the intriguing possibility that the acetylation of K23 may serve to inhibit cleavage at certain sites.

Cathepsin L Is Associated with Chromatin *In Vivo*

Although there is previous evidence that Cathepsin L and its activity exists in the cell nucleus of MEFs (Goulet et al., 2004), I wanted to document this nuclear localization in our ESC model using a biochemical approach: I isolated chromatin and then solubilized it by micrococcal nuclease digestion as described above (**Figure 4.1C**). As shown by probing with Cathepsin L antibody, Cathepsin L was indeed associated with chromatin fragments released by nuclease digestion (**Figure 4.8D**). Importantly, chromatin association only begins to appear upon differentiation and is not apparent in undifferentiated ESCs. I note as well that the mature form of Cathepsin L (marked by asterisk) was enriched over the full length (hash) or intermediate (dot) forms.

rCathepsin L Reproduces the *In Vivo* Histone H3 Cleavage Pattern

To further validate the identification of Cathepsin L as an H3 protease, I tested recombinant mouse Cathepsin L enzyme in our H3 cleavage assay (**Figure 4.9**). Since Cathepsin L displays optimal activity at the more acidic pH of the lysosome (i.e. ~pH 5, (Barrett et al., 2003)), and since I suspected that Cathepsin L is acting within the nucleus at a higher pH (~pH 7-8), I tested the recombinant enzyme in both pH environments. Importantly, recombinant mouse Cathepsin L was able to cleave recombinant histone H3 at both pH 5.5 and pH 7.4. Moreover, the recombinant enzyme cleaved H3 and created the specific epitope recognized by the cleavage site-specific H3.cs1 antibody (**Figure 4.9A**, lower panel). To further characterize the sites of proteolytic cleavage, I sent the reaction products of rCathepsin L and rH3-HIS to the Hunt Laboratory for analysis by MS. Importantly, MS analysis revealed that the six sites of cleavage mapped in the endogenous H3 samples (**Figures 4.3 and 4.4**) were also produced by rCathepsin L *in vitro* (**Figure 4.9B**). Also observed in both reactions are five of the six complementary N-terminal peptides generated by rCathepsin L cleavage of H3 (left), and their abundances correlated with that of their C-terminal counterparts. Furthermore, the most abundant site of cleavage *in vitro* was between A21 and T22, the same site that was shown to be most abundant in the MS analysis of endogenous histone H3. In addition, the relative abundances of the *in vivo* H3 cleavage sites and those produced by Cathepsin L *in vitro* are highly correlative. One exception is cleavage between residues K27 and S28, which is more abundant in the *in vitro* assay compared to that mapped from *in vivo* samples (the latter shows a preference for cleavage between residues R26/K27 over K27/S28). We speculate that the presence of histone modifications *in vivo* play a role in regulating the preferred cleavage site of Cathepsin L, particularly at sites surrounding H3K27, which was shown to be preferentially methylated on cleaved H3 by both immunoblot and MS analysis (**Figure 4.1A and Figure 4.4**).



B

MS analysis of rCathL cleavage

N-terminal peptides		C-terminal peptides	
1-21	ARTKQTARKSTGGKAPRKQLA	TKAARKSAPATGGVKK…	22-50
1-22	ARTKQTARKSTGGKAPRKQLAT	KAARKSAPATGGVKK…	23-50
1-23	ARTKQTARKSTGGKAPRKQLATK	AARKSAPATGGVKK…	24-50
1-24	n.d.	ARKSAPATGGVKK…	25-50
1-25	n.d.	n.d.	26-50
1-26	ARTKQTARKSTGGKAPRKQLATKAAR	KSAPATGGVKK…	27-50
1-27	ARTKQTARKSTGGKAPRKQLATKAARK	SAPATGGVKK…	28-50
1-36	ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVK	K...	37-50

data in B produced by Tara Muratore-Schroeder and used with permission

Figure 4.9: rCathepsin L cleaves histone H3 *in vitro*.

A. Recombinant Cathepsin L cleaves rH3 *in vitro* at both pH 5.5 and pH 7.4 and generates a fragment that is recognized by both α-HIS-HRP (top panel) and α-H3.cs1 (middle panel) antibodies.

B. Recombinant mouse Cathepsin L was incubated with recombinant H3-HIS at both pH 5.5 and pH 7.4; after 2 hours, the reaction products were subjected to analysis by mass spectrometry. Both N-terminal (left) and C-terminal (right) fragments of the rH3 cleavage were detected; note the similarity to the pattern of *in vivo* cleavage shown in **Figure 4.3C**. Summed ion currents for all charge states of the unmodified sequences were employed to estimate the relative abundances of the six peptides beginning with the indicated residues as follows: T22 > K23, A24, and S28 > A25 and K27. Also observed in both of the above digests are five of the six complementary N-terminal peptides generated by Cathepsin L cleavage of H3 (left panel) containing residues A1-A21, A1-T22, A1-K23, A1-R26 and A1-K27. Relative abundances of these N-terminal peptides correlate, as expected, with their C-terminal counterparts: A1-A21 > A1-T22, A1-K23, A1-K27 > A1-R26.

C. Recombinant Cathepsins B, K, and L were pre-activated and incubated with recombinant H3-HIS at both pH 5.5 and pH 7.4; after 15 minutes, the reactions were separated by SDS-PAGE and analyzed by immunoblotting.

Taken together, the above data support the conclusion that Cathepsin L is capable of generating all of the histone H3.2 fragments observed at day three following induction of ESC differentiation by retinoic acid.

I then wanted to test other cathepsin family members in the H3 cleavage assay. I chose to test rCathepsin B, one of the most abundant lysosomal proteases, and rCathepsin K, which is reported to have a significant preference for leucine in

the P2 position of its substrates much like Cathepsin L (Cathepsin S also shares this preference but was not found to have significant expression in our ESC model). Following pre-activation the recombinant enzymes, Cathepsins B, K and L were incubated with rH3-HIS at both pH 5.5. and pH 7.5, and the reactions were then analyzed by immunoblotting with both HIS and H3.cs1 antibodies. As shown in **Figure 4.9C**, pre-activated Cathepsin L cleaves rH3-HIS robustly and produces a pattern that is similar to that observed *in vivo.* In contrast, Cathepsin B cleaves rH3 with a distinct pattern from Cathepsins K or L and does not significantly produce the H3.cs1 epitope. Cathepsin K produces a similar pattern to Cathepsin L under these assay conditions, although they appear to differ in their preferences for specific sites (e.g. the cleavage site recognized by the H3cs.1 antibody is less abundant in the Cathepsin K reactions).

Both RNAi-mediated and Chemical Inhibition of Cathepsin L Inhibits H3 Cleavage *In Vivo*

I then sought to show that knockdown of Cathepsin L *in vivo* led to a decrease in H3 cleavage. I first created heterogeneous (non-clonal) stable cells lines constitutively expressing short RNAi hairpins to both a control gene (a human gene whose hairpin sequence show no significant sequence homology to the mouse genome by BLAST) and the Cathepsin L gene (*Ctsl*) by puromycin selection. I then differentiated these heterogeneous cell lines with RA, as usual, and harvested the cells at various time points. As shown in **Figure 4.10A and B**, knockdown of Cathepsin L (shown by immunoblotting with Cathepsin L antibody, upper panels) led to reproducible decrease in H3 cleavage as detected by the H3 C-terminal antibody (lower panels). In parallel, samples taken at day 3 post-induction with RA were also evaluated for knockdown efficiency and H3 cleavage by titration of sample (**Figure 4.10B**).

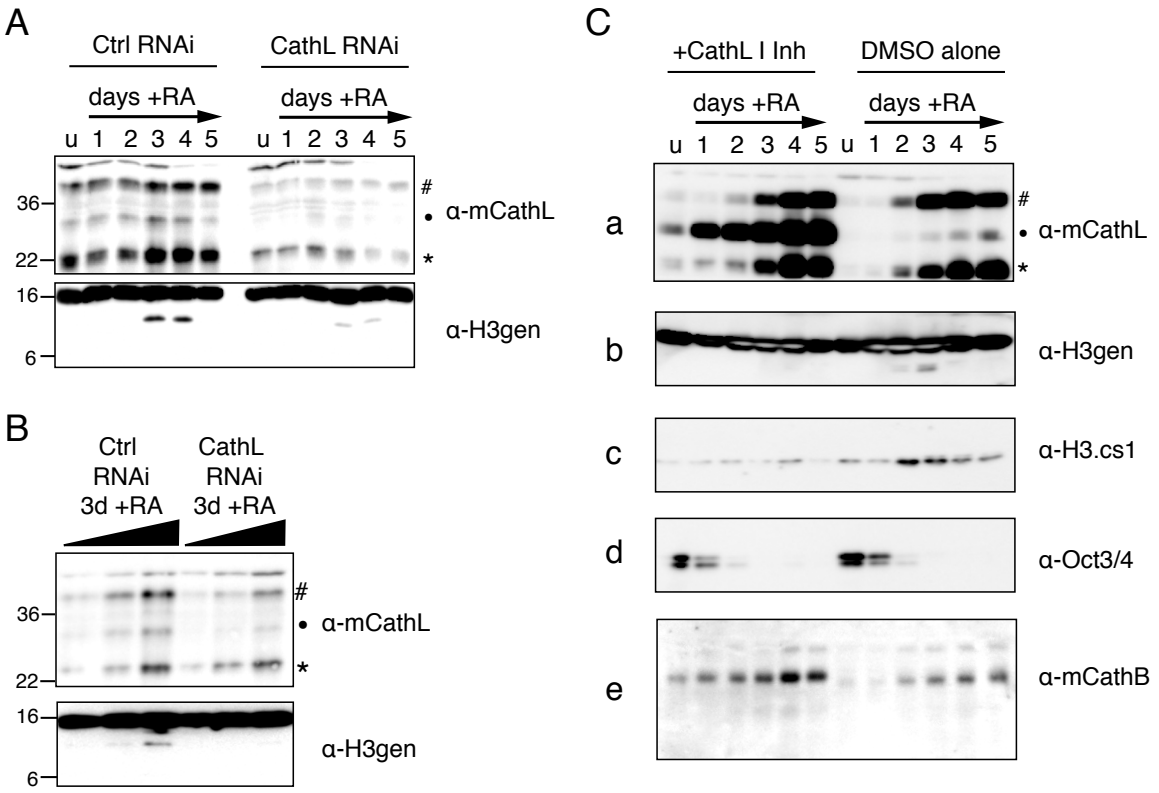


Figure 4.10: Both RNAi and chemical inhibition of Cathepsin L reduce histone H3 cleavage *in vivo*.

A. Control and *Ctsl* RNAi cells lines were differentiated with RA as usual and harvested at the indicated time points; WCEs were then separated by SDS-PAGE and analyzed for both Cathepsin L expression (upper panel) and histone H3 cleavage (lower panel) by immunoblotting.

B. A serial two-fold dilution of samples from day 3 +RA induction were resolved by SDS-PAGE gel and analyzed by immunoblotting as in **A**.

C. The addition of Cathepsin L Inhibitor I to the cell media of differentiating ESCs (left side) inhibits the processing of Cathepsin L itself (a) as well as that of histone H3 (b, c) as compared to DMSO alone treated control cells (right side). Loss of pluripotency marker Oct 3/4 was not affected (d) nor was the self-processing of another cathepsin family member, Cathepsin B (e).

To further assess whether Cathepsin L causes histone H3 cleavage *in vivo*, I treated differentiating ESCs with a commercially available, cell-permeable Cathepsin L specific inhibitor, Cathepsin L Inhibitor I. Inhibition of Cathepsin L using a chemical inhibitor allowed me to assess the effect of Cathepsin L inhibition within a single, wild-type cell line rather than comparing individual cell lines created by drug selection. Undifferentiated ESCs were treated with or without inhibitor for 24 hours prior to plating for differentiation. Cells were then differentiated with RA while either Cathepsin L Inhibitor I or DMSO alone was maintained in the media.

As shown in **Figure 4.10C**, cells treated with inhibitor showed significant failure to fully process Cathepsin L from its intermediate form (marked by dot) into its mature form (asterisk), demonstrating that its self-activating activity had been inhibited (**Figure 4.10C**, panel a, left). In contrast, no significant accumulation of the intermediate form was seen in the DMSO alone treated cells (**Figure 4.10C**, panel a, right). Notably, histone H3 cleavage also decreased significantly in the cells treated with the inhibitor, but not in cells treated with DMSO alone, as shown by probing immunoblots with both H3-general and H3.cs1 antibodies (**Figure 4.10C**, panels b and c). Interestingly, all cleavage detected by the H3-general antibody was abolished in the cells treated with inhibitor, suggesting that Cathepsin L may be responsible for all sites of cleavage (supported by MS data in **Figure 4.9**).

Importantly, neither Oct-3/4 nor Cathepsin B levels (**Figure 4.10C**, panels d and e) appeared to change upon addition of the inhibitor. Since Oct-3/4 is a marker of pluripotency that is normally lost rapidly upon differentiation with RA (**Figure 4.1D**), the fact that this pattern is unchanged in the presence of the Cathepsin L Inhibitor I indicates that the inhibitor does not alter the pluripotency or differentiation capacity of ESCs prior to the time point at which H3 cleavage is observed. This conclusion is supported by Q-PCR data for the pluripotency marker Nanog (**Figure 4.11A**), which also remains unaffected by Cathepsin L inhibition. The consistency of Cathepsin B levels between those cells treated with Cathepsin L Inhibitor I and control cells suggests that the inhibitor is indeed specific for Cathepsin L. Markers of cell lineage were also analyzed by Q-PCR and suggest some changes in neural/ ectodermal expression patterns and levels of endodermal marker expression between inhibitor treated and control cells (**Figure 4.11B and C**). However, the changes observed were not statistically significant and, in fact, add more support to the conclusion that Cathepsin L I inhibitor does not globally affect the pluripotency or differentiation profile of ESCs. Moreover, given that the inhibitor (or the RNAi

Figure 4.11: The effect of *in vivo* chemical inhibition of Cathepsin L on ESC differentiation.
A. Gene expression of pluripotency marker Nanog: cells were differentiated with RA in the presence of Cathepsin L I inhibitor or DMSO alone, as in **Figure 4.10C**; mRNA was then isolated, reverse transcribed into cDNA and assayed by Q-PCR with SYBRgreen. Data in panels **A**-**C** represent the mean ±SD of three independent experiments.
B. Q-PCR for Nestin, a neuronal/ectodermal marker, suggests its expression may increase with Cathepsin L I inhibitor on days 3, 4 and 5 of differentiation versus control cells. Other neuronal markers did not show this trend (Pax3, Musashi, HoxA1).
C. Q-PCR for Sox17, a marker of endoderm, showed some difference between inhibitor treated and control cells in level of expression (expression was higher in control cells, $P \approx 0.5$ for days 4 and 5); another endodermal marker, GATA6, did not show significant difference when averaged over three experiments. Interestingly, Q-PCR for markers of mesodermal lineage, including Myf5 and Fgf8, did not detect significant expression.

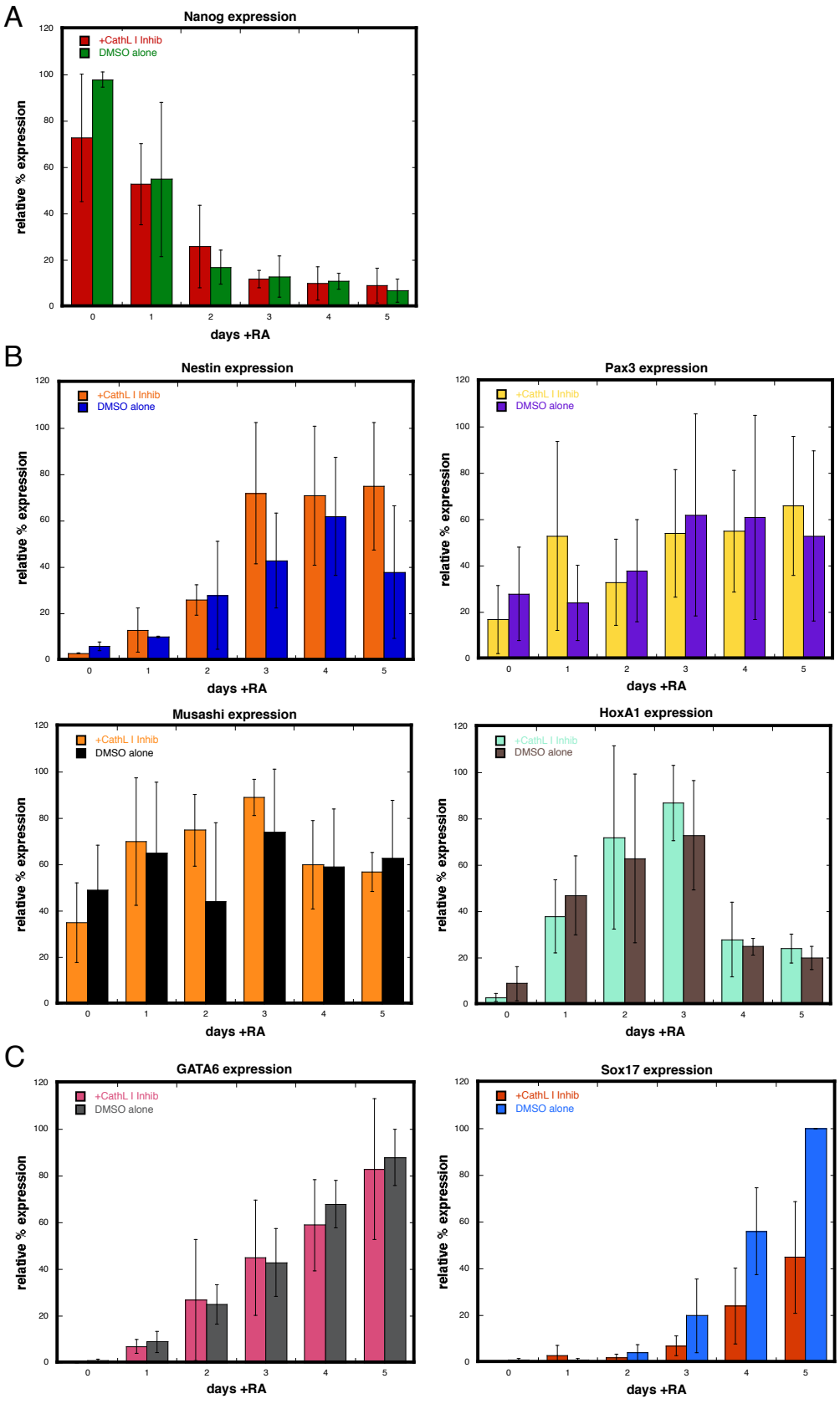


Figure 4.11: The effect of *in vivo* chemical inhibition of Cathepsin L on ESC differentiation.

hairpin) is not specific to the nuclear Cathepsin L enzyme and that the differentiation of ESCs with RA leads to a heterogenous population of cell types (despite being less heterogeneous than other differentiation methods), this type of experiment is unlikely to reveal the specific effects of H3 cleavage inhibition.

Histone Tail Modifications Can Modulate Cathepsin L Activity and Its Downstream Effects

I then asked whether known covalent modifications on the histone tail affect the cleavage activity of Cathepsin L. As depicted in **Figure 4.12A**, several amino acids near the sites of Cathepsin L cleavage are known to be modified by either acetylation (triangles) or methylation (circles). To test the potential effects of these modifications on the cleavage activity of Cathepsin L, I first turned to the H3 cleavage assay described in **Figure 4.5**. Four different recombinant H3 substrates were prepared as follows: unmodified rH3 (1), rH3 dimethylated specifically at K27me2 (2), rH3 "pan-acetylated" by treatment with acetic anhydride (3) and rH3 with both K27me2 and pan-acetylation (4). These substrates were shown to have the specific modifications of interest depicted in **Figure 4.12A** by immunoblot (**Figure 4.12B**) and were verified by MS (≥ 90%, data not shown). As shown in **Figure 4.12A**, this assay demonstrated that acetylation of lysine residues greatly reduced cleavage of H3 by rCathepsin L at both pH 7.5 and 5.5 (compare substrate 1 to 3). In contrast, K27me2 increases H3 cleavage (compare substrate 1 to 2; greater depletion of full-length rH3+K27me2 suggesting increased cleavage activity at pH 5.5).

In order to test the effect of the above modifications depicted in **Figure 4.12A** more quantitatively and specifically, I synthesized a set of five peptides with identical backbone sequence flanking the H3 cleavage site (H3 15-31) modified as follows: unmodified, K18 acetyl, K23 acetyl, K18ac+K23ac, and K27me2. I then

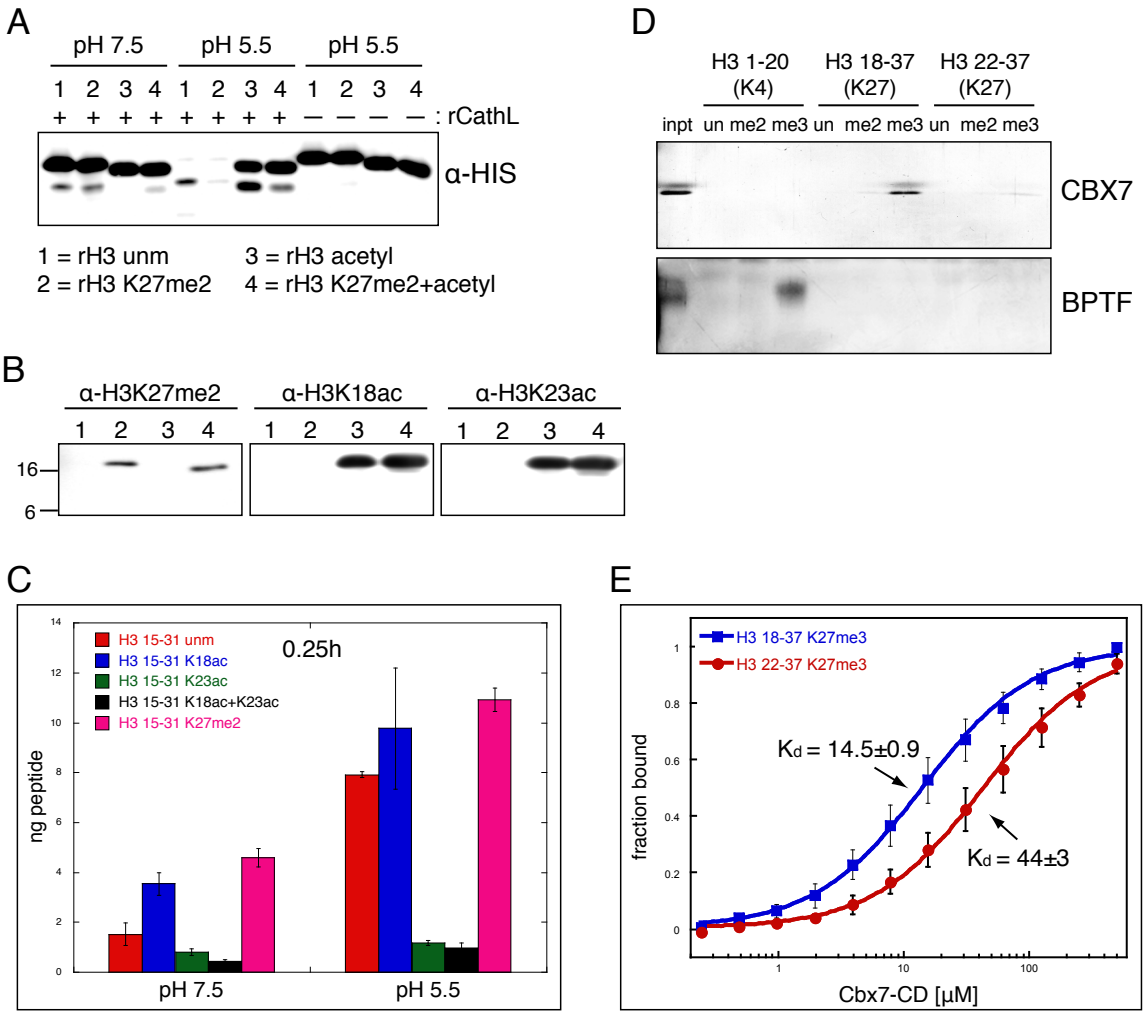


Figure 4.12: Covalent histone modifications modulate Cathepsin L activity and its downstream effects.

A. rCathepsin L was incubated with each of four rH3 substrates: 1= rH3 unmodified, 2= rH3 alkylated to K27me2, 3= rH3 pan-acetylated, 4= rH3+K27me2 pan-acetylated.

B. C-terminally HIS-tagged recombinant histone H3 was mutated to K27C and then alkylated to convert K27C to K27me2; an aliquot of this protein was then treated with acetic anhydride to acetylate all free lysines. These proteins were then assayed for the presence of K27me2, H3K18ac and H3K23ac by immunoblot..

C. H3 cleavage reactions were performed as in **A** using synthesized peptides that represent the H3 tail from amino acids 15 to 31. Reactions were incubated with ~250pmol peptide and quenched with 0.1% TFA before being plated in duplicate for ELISA with the H3cs.1 antibody. Signal was normalized to that of mock reactions for each peptide. Results represent the mean of three independent experiments ± standard deviation (SD).

D. Peptide pull-down assays were performed with recombinant mouse CBX7-chromodomain or rhBPTF-PHD finger and H3 peptides (sequences and methylation states indicated).

E. Fluorescence anisotropy Cbx7-CD protein binding to non-cleaved peptide (18-37) vs. cleaved peptide (22-37); binding decreases 3 fold with cleaved peptide, $p<0.01$. K_ds are in μM±SEM. Data points represent the mean±SD.

employed an ELISA based assay using the H3cs.1 antibody described previously (**Figure 4.6**) to quantitatively measure the H3cs.1 cleavage activity of rCathepsin L on these peptides. As shown in **Figure 4.12C**, clear differences in rCathepsin L activity were observed at both pH 7.5 and pH 5.5. As suggested by the rH3 cleavage assay described above, K27me2 (magenta) strongly increases the ability of Cathepsin L to cleave at H3cs.1 as compared to the matched unmodified H3 15-31 peptide (red). Interestingly, acetylation at K18 (blue) also increases this activity, suggesting that the acetylation of another lysine or combination of lysines must be responsible for the abrogation of cleavage by acetylation demonstrated above. Our data suggest that acetylation at K23ac is at least partly responsible for this effect, as the K23ac peptide shows very little cleavage activity at H3cs.1, both alone (green) and in combination with K18ac (black).

I then asked what effect of the cleavage of the H3 tail might have on the binding of effector proteins. Since K27 methylation is a well-documented binding site for the chromodomain-containing protein Polycomb (Pc) (Bernstein et al., 2006b; Fischle et al., 2003), I sought to test the effect of H3 cleavage on Pc binding to H3K27 methyl. Using peptides that mimicked either the “non-cleaved” H3 tail (H3 18-37) or the “cleaved” H3 tail (H3 22-37), I assayed the effects of H3 cleavage on the ability for the chromodomain of Pc (mouse CBX7) to bind to K27 methyl (**Figure 4.12D**); the PHD-finger of the known H3K4 methyl binding protein BPTF was used as a positive control and to demonstrate specificity (**Figure 4.12D, lower panel**). As shown in **Figure 4.12D** and quantified in **Figure 4.12E**, H3 cleavage greatly diminished the ability of CBX7 to bind to H3K27 methyl, suggesting that proteolytic modification introduced by Cathepsin L could lead to significant downstream effects.

Chapter 4 Discussion

Here I show that ESCs employ regulated histone proteolysis in order to change their "epigenetic signature" upon differentiation. Furthermore, I have identified Cathepsin L as a protease responsible for developmentally-regulated histone H3 cleavage and demonstrated that its activity may be modulated by the modification of the histone tail itself. This study reveals a novel nuclear function of this family of cysteine proteases in histone and stem cell biology and suggests that controlled histone proteolysis may be part of a mechanism for introducing variation into the chromatin polymer.

Despite remarkable progress made in documenting epigenetic signatures such as "bivalent domains" i.e. an H3 tail bearing both H3K4me3 and H3K27me3 marks (Bernstein et al., 2006a), little is known as to what mechanisms may function to bring about the above changes at a chromatin level. Our data suggest that mouse ESCs employ a novel, regulated histone H3 proteolysis mechanism that may serve to alter epigenetic signatures upon differentiation. Other means of actively removing histone methyl marks have been well documented, including enzymatic demethylation (Anand and Marmorstein, 2007; Shi et al., 2004) and selective histone variant replacement (Ahmad and Henikoff, 2002). However, until now considerably less evidence of endogenous histone proteolysis has been reported in mammalian cells, although N-terminal cleavage of the histone H3 tail has been shown in a ciliate model, *Tetrahymena* (Allis et al., 1980), in mammalian cells infected with the foot-and-mouth disease virus (Falk et al., 1990), and recently in the budding yeast *s.cerevisiae* (Santos-Rosa et al., 2009).

The finding that Cathepsin L cleaves histone H3 in mouse ES cells was unexpected as this enzyme was originally described as a lysosomal protease (Barrett et al., 2003). However, Cathepsin L has been shown to localize to nuclei

where it plays a role in the proteolytic processing of transcription factor CDP/Cux (Goulet et al., 2004). Notably, my data show that the activities of both fraction 23 and recombinant Cathepsin L are higher at pH 5.5 than at pH 7.4 (**Figure 4.9**); however, as noted in Goulet et al, this may in fact be an important feature in the regulation of its activity in the nucleus. Biochemical studies have also shown that pro-Cathepsin L is localized to the nucleus in *ras*-transformed mouse fibroblasts (Hiwasa and Sakiyama, 1996). In addition, evidence of endogenous, nuclear serpin inhibitors with the ability to inhibit Cathepsin L activity, e.g. MENT (Bulynko et al., 2006) and Cystatin B (Riccio et al., 2001), further supports the notion that Cathepsin L, and potentially other cysteine proteases, play important but poorly understood roles in regulated nuclear proteolysis. As far as I am aware, transcription factor CDP/Cux is the only nuclear substrate of Cathepsin L identified to date, and histones have not yet been identified as substrates of this class of proteases in mammalian cells. Interestingly, however, recent studies in sea urchin have suggested that a Cathepsin L-like cysteine protease may be responsible for the degradation of sperm histones during a key chromatin remodeling event after fertlization (Morin et al., 2008). My data support other studies that demonstrate the nuclear localization of Cathepsin L and provide the first indication that cellular histones, H3 in particular, are key substrates of this family of proteases in mammalian cells.

The finding that Cathepsin L is an H3 protease is interesting when considering the phenotype common to the Cathepsin L knockout mouse (Nakagawa et al., 1998) and the *furless* mouse, which has been shown to have a spontaneous mutation in the Cathepsin L gene (Roth et al., 2000). These mice exhibit periodic hair loss due to the improper cycling and morphogenesis of their hair follicles. Roth et al conclude that upon the loss or mutation of Cathepsin L “the balance between proliferation and differentiation seems to be shifted toward proliferation,” and subsequently causes the observed phenotype. Since skin cells derive from

an ectodermal lineage and also maintain a degree of “stemness” in their ability to both proliferate and differentiate, it is interesting to speculate that perhaps the role of Cathepsin L in histone H3 cleavage is repeated again in other models of differentiation. Cathepsin L knockout mice are viable and fertile, however, indicating that its functions are nonessential and/or redundant. That said, these mice have been reported to show defects in spermatogenesis (Wright et al., 2003), which is interesting in light of the above mentioned cathepsin-mediated degradation of sea urchin sperm histones.

Notably, Cathepsin L/Cathepsin B double knockout mice exhibit severe brain atrophy and die two to four weeks after birth (Felbor et al., 2002). The severity and selectivity of this phenotype suggests that these two enzymes overlap in their specific functions. Although I did not observe significant inhibition of Cathepsin B upon *in vivo* chemical inhibition of Cathepsin L, nor did I observe the same pattern of H3 cleavage with recombinant Cathepsin B compared to Cathepsin L *in vitro*, I cannot exclude the possibility that redundancy in H3 cleavage function may exist between these or other related enzymes at other stages or lineages of differentiation. One potentially interesting source of redundancy might be a poorly studied family of cathepsins known as the placental cathepsins (Barrett et al., 2003). In rodents, a series of tandem duplications of the Cathepsin L gene (*ctsl*) has produced a family of eight additional cathepsin genes on chromosome 13. Originally cloned from the placenta of rat and mice (Conliffe et al., 1995; Sol-Church et al., 2000a; Sol-Church et al., 2000b, c), these Cathepsin L-related genes are differentially expressed and may have distinct roles both within the placenta and other tissues (Barrett et al., 2003).

These findings require a revaluation as to the function of this family of enzymes in transcriptional and epigenetic regulation; many important questions remain. First, how is the protease cleavage activity regulated? Along this line,

I showed that covalent modifications (i.e. acetylation and methylation of nearby lysines) serve to regulate, positively and negatively, the H3 proteolytic processing event (Figure 27). Also note that proteolytic processing of H3 in the ciliate model occurs selectively in a *hypo*acetylated, transcriptionally silent (micronuclear) genome, while processing of H3 fails to occur in *hyper*acetylated, transcriptionally active macronuclei (Allis et al., 1980). These data raise the question as to whether other chromatin-modifying enzyme complexes, such as HATs, HDACs, or ubiquitinating enzymes, play a role in regulating histone proteolysis.

Second, by what mechanism is the cleaved H3 replaced, and is DNA replication and chromatin assembly required? Although it is possible that the cell would leave those loci in a permanently changed state, one would imagine that the cell would have a mechanism to reverse the effect of cleavage. My data supports this in that I do not observe the cleaved H3 product in ESCs beyond 4 days of RA induction. In fact, the cell has a few different mechanisms that might resolve this issue. For example, since not every H3 molecule will have been cleaved (or even serve as a substrate for the enzyme), the cell could use remodeling mechanisms to remove the cleaved H3 nucleosomes and replace it with those from other regions of the genome. Additionally, the proteolyzed molecule of H3 may be removed from the chromatin and replaced by a full-length H3 molecule, either by any H3 isoform during S phase or by H3.3 outside of S phase (Loyola and Almouzni, 2007). As I described in my work with Sandra Hake in Chapter 3, the ratios between H3 isoforms change as ESCs differentiate, suggesting that H3 replacement mechanisms in ESCs are highly robust and could rapidly replace a tail-less H3 molecule. My findings with the Hunt laboratory also show that the histone variant H3.2 is preferably cleaved as compared to H3.3, supporting the idea that S-phase/replication-coupled replacement may be involved. Along this line, I also note that proteolytic processing of transcription factor CDP/Cux by

Cathepsin L occurs during the G1/S-phase transition and is coupled to cell cycle progression (Goulet et al., 2004). Whether proteolytic processing of H3 is cell cycle dependent remains unclear, although it is interesting to note that my studies were done in ESCs, which cycle rapidly and spend a high percentage of time in S-phase as measured by flow cytometry (**Figure 3.2D**).

Third, and more broadly, does this proteolytic mechanism apply to other histone substrates or to other cathepsin family members or cathepsin-like proteases, and is the general mechanism conserved in other organisms or operating in other stages of development or differentiation? Although Cathepsin L is the only member of this family shown to localize to the nucleus, the conservation in genomic structure and sequence between this protease and its related family members suggest that other cathepsins may function similarly. As with other chromatin-modifying enzymes/complexes, I suspect that the particular enzyme system and mechanism may not be universal, but rather specific to a particular developmental time and/or genomic locus. Moreover, it is possible, if not likely, that histone H3 is only a substrate of Cathepsin L during a particular window of time and space, dictated by multiple modes of regulation that may include the presence of endogenous protease inhibitors (e.g. MENT (Bulynko et al., 2006)) and the modification of the H3 tail itself.

Finally, the major outstanding question is: what is the downstream biological function of histone H3 tail cleavage? In order to assay for the downstream effects of H3 proteolysis, this event must be separated from the effects of its cleavage of other Cathepsin L substrates. Experimentally, this is quite difficult, since there is at least one other known nuclear substrate, transcription factor CDP/Cux (Goulet et al., 2004). One possibility would be if the structure of mammalian Cathepsin L + H3 peptide revealed unique interactions that would then allow for the over-expression of a mutant protein that only affected its activity toward histone H3.

Another possibility would be to replace all copies of H3 with a non-cleavable protein, such as one of the mutants in **Figure 4.8**; however, given that mammalian cells have many (>10) copies of the histone H3 gene in their genome, this experiment may not be feasible. It would also be interesting to test the hypothesis that H3 proteolysis leads to the regulated replacement of that histone at specific genomic loci. In order to do this, it would be necessary to develop ChIP methods to assay the removal of cleaved H3 and its replacement (e.g. w/H3.3). Although I did make several attempt to ChIP differentiating ESC chromatin with the H3cs.1 antibody I developed, I never found significant enrichment at specific genes. It might be necessary to generate new antibodies and screen for ones that are ChIP-grade. Once a map is generated of the loci at which H3 is cleaved, it would then be possible to develop ChIP approaches to detect its replacement. Another obstacle to this experiment will be the lack of H3 antibodies that can distinguish between the three H3 isoforms. However, once the genomic targets of H3 cleavage are known it would be possible to try using epitope tagged H3 isoforms to analyze their presence at these loci before and after H3 cleavage.

Limited proteolysis of nuclear proteins is an important means of regulating transcription and other cell processes (Goulet and Nepveu, 2004; Vogel and Kristie, 2006). Here I show that limited proteolysis of histone H3 by Cathepsin L occurs during the differentiation of ESCs and may be regulated both positively and negatively by covalent modifications on the H3 tail itself. Despite the technical difficulties in assaying its function, the identification and characterization of H3 cleavage by Cathepsin L during ESC differentiation is an important step in understanding limited nuclear histone proteolysis as a mode of transcriptional regulation. It is possible that H3 cleavage serves to remodel chromatin and/or regulate gene expression, perhaps by removing certain covalent modifications, and/or creating a new N-terminal H3 tail (as well as a new sequence context for

binding modules), and/or altering the structure of selected nucleosomes (affecting their compaction and stability). I will address these questions and other possible far-reaching implications of regulated histone proteolysis in the General Discussion in Chapter 5.

Chapter 5: General Discussion

A rapidly growing literature demonstrates that cells undergo dramatic developmentally-regulated changes in gene expression and cellular morphology upon transition from undifferentiated stem cells to specific lineages (Giadrossi et al., 2007; Kim et al., 2008; Murry and Keller, 2008). Chromatin proteins provide cells with layers of gene regulation that are independent of genomic sequence. Here I have described just a few of the many chromatin mechanisms that cells use to control differentiation and development. In particular, my work focuses on changes that involve histone H3 during the differentiation of mouse Embryonic Stem Cells (ESCs). Below I will discuss some of the potential far-reaching implications of this work. I will focus mainly on the study of H3 proteolysis during ESC differentiation (described in Chapter 4), since this work was the major focus of my graduate studies; however, I will discuss how other aspects of H3 regulation during ESC differentiation, including histone replacement and the recruitment of specific histone-binding proteins, may converge into common regulatory mechanisms.

There is now growing evidence supporting the idea that histones, particularly histone H3, are proteolytically cleaved in the course of normal growth and development. Past studies cell-cycle documented the cell-cycle regulated, nuclear cleavage of H3 in the ciliate *Tetrahymena* (Allis et al., 1980) and the proteolysis of H3 in mammalian cells infected with the foot-in-mouth-disease virus (FMDV) (Falk et al., 1990). More recent work includes the study described here (in Chapter 4) demonstrating H3 cleavage in differentiating mouse ESCs (Duncan et al., 2008) and a report of N-terminal H3 proteolysis in budding yeast (Santos-Rosa et al., 2009). Interestingly, the major site of cleavage is conserved in these last two studies (between alanine 21 and threonine 22), suggesting that such regulated proteolysis may be evolutionarily conserved. Here, I show that H3 cleavage in differentiating

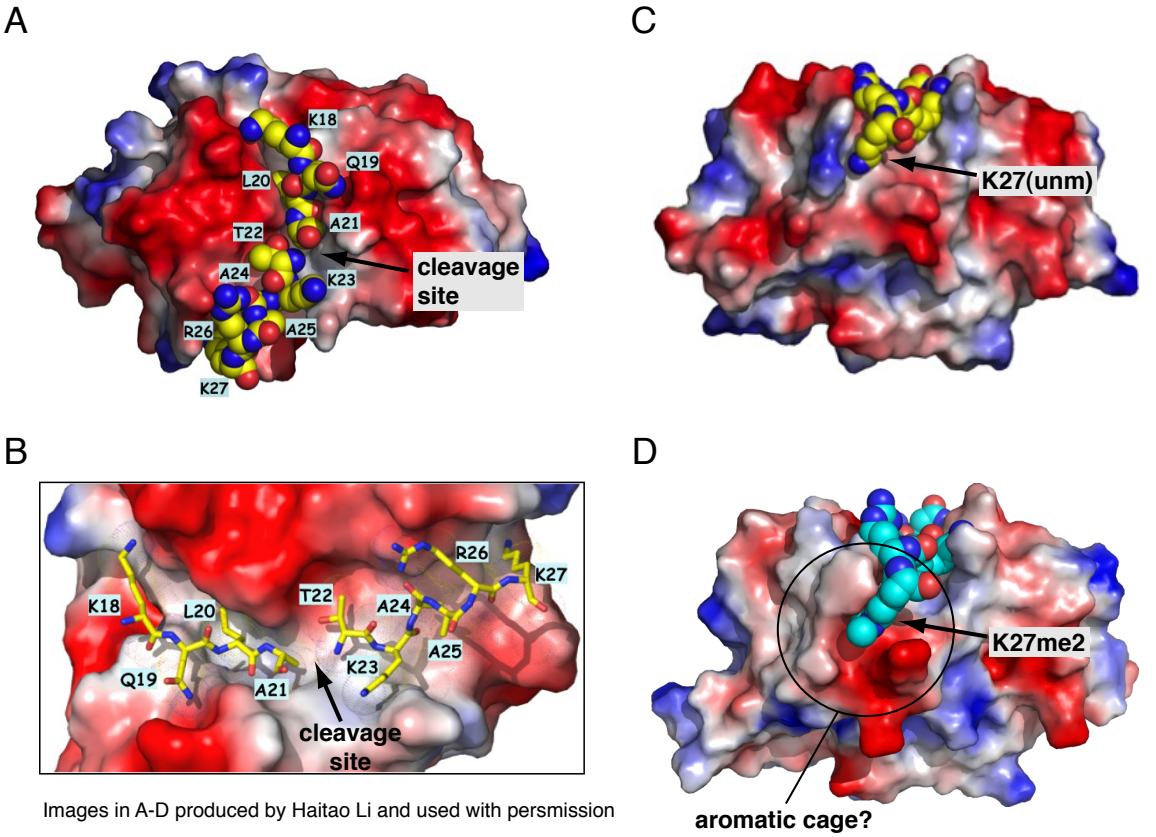
ESCs can be performed by the papain-like cysteine protease Cathepsin L and that its activity toward H3 is affected by modifications on the histone tail itself. Nevertheless, many questions remain regarding the regulation and mechanism of this particular H3 proteolysis and, more broadly, the potential functions and downstream effects of regulated histone proteolysis in general. Below I will discuss some of the possible mechanisms and functions of regulated histone proteolysis and how they may cooperate with other mechanisms of chromatin remodeling to affect epigenetic regulation.

Substrate Recognition by Cathepsin L

One outstanding question about the regulated H3 proteolysis observed in mouse ESCs concerns the selective cleavage of the H3 tails. Three observations described in Chapter 4 suggest that H3 cleavage in differentiating ESCs occurs preferentially on a subset of H3 tails: 1) only a small percentage of H3 is cleaved, 2) H3 variant H3.2 is preferentially cleaved over H3.3 and possibly H3.1, and 3) the cleaved H3 is enriched in certain histone modifications, particularly H3K27me2. Although it is possible that other proteins mediate this preference as part of a protease complex, it is also feasible that the enzyme itself, specifically Cathepsin L, directs this selectivity. One way to investigate this possibility is through a structural analysis of the Cathepsin L enzyme in complex with H3 peptides, in the same way that the structural studies described above on the PHD finger of BPTF revealed the nature of its ability to bind H3K4me3. Dr. Patel and colleagues (MSKCC) are currently trying to obtain crystal structures of Cathepsin L + H3 peptide complexes; however, they have successfully obtained crystal structures of the cathepsin enzyme alone and used this data to model its recognition of H3 peptides. As shown in **Figure 5.1**, these models not only strongly support other data showing that Cathepsin L cleaves histone H3 between alanine 21 and threonine 22 (**Figure 5.1A and B**),

but also suggest that the enzyme may have an aromatic cage, similar to those described above found in chromodomains (CDs) and PHD fingers, that could engage K27me2 if it assumes certain conformations (**Figure 5.1C and D**).

These exciting structural studies could play an important role in forming hypotheses and guiding future functional experiments. Importantly, they will help to better define not only how H3 is cleaved at specific sites, but also how it is recognized by Cathepsin L and potentially other papain-like cysteine proteases.



Images in A-D produced by Haitao Li and used with persmission

Figure 5.1: Cathepsin L may recognize dimethylated H3K27.

A. Electrostatic surface representation (red = negatively charged and blue = positively charged surface) of the crystal structure of Cathepsin L with an unmodified H3 (aa 18-27) peptide modeled into the complex; the peptide is shown in a space filling representation. The black arrow indicates the cleavage site.

B. An expanded view of the cleavage site in **A**.

C. The model in **A** rotated 90 degrees to show the accommodation of the unmodified K27.

D. The structure shown in **A** with an H3 (aa 18-27) K27me2 peptide modeled into the complex; the black arrow indicates the dimethyl K27 and the circle highlights the potential "cage" that accommodates the methylammonium group. (Images in **A**-**D** produced by Haitao Li and used with permission.)

Such information could potentially lead to the discovery of new small-molecule inhibitors (or activators) that perturb this activity with greater specificity than those currently available. It could also pinpoint the structural features critical to H3 recognition, which could guide mutagenesis studies that aim to dissect the function of Cathepsin L activity toward histone H3 versus its other nuclear substrate(s).

Histone Proteolysis and Chromatin Remodeling

Another outstanding question about H3 proteolysis during ESC differentiation, and histone proteolysis in general, is: what are the downstream effects? Although this question is still largely unanswered, data presented here and in other studies (Allis et al., 1980; Santos-Rosa et al., 2009) offer some hints and possibilities.

As shown in **Figure 5.2**, H3 cleavage could impact chromatin function at the level of the histone tail (**Figure 5.2A**), the nucleosome, or the organization of the chromatin fiber (**Figure 5.2B**). In Chapter 4, I showed evidence that cleavage of the H3 tail by Cathepsin L significantly decreases the binding affinity of mammalian Polycomb protein CBX7 for H3K27me3. This effect is not surprising since structural studies of the CD of *Drosophila* Polycomb in complex with an H3 peptide (Min et al., 2003) show that it makes contacts with amino acids N-terminal to the cleavage site mapped in differentiating ESCs (represented by the star shape in **Figure 5.2A-1**). It is also the case that Polycomb (and mammalian Pc CBX7) has a higher affinity for H3 trimethylated at K27, while the cleaved H3 is preferentially dimethylated at K27. That said, CBX7 does bind with some affinity to H3K27me2, and it is unclear whether *in vitro* binding affinities directly translate to *in vivo* preferences for methylation state. In any case, this evidence supports the idea that histone cleavage can impact downstream cellular function by affecting the binding of effector proteins such as Polycomb.

The fact that histone cleavage can impact the affinity of known histone methyl-binding proteins also raises the question as to whether such cleavage could also lead to the recruitment of new effector proteins. Given that the CD of Polycomb and the binding domains of other chromatin-interacting proteins (e.g. the PHD finger of BPTF) are known to contact and recognize both the methyl modification on a particular lysine and the backbone sequence surrounding it, as

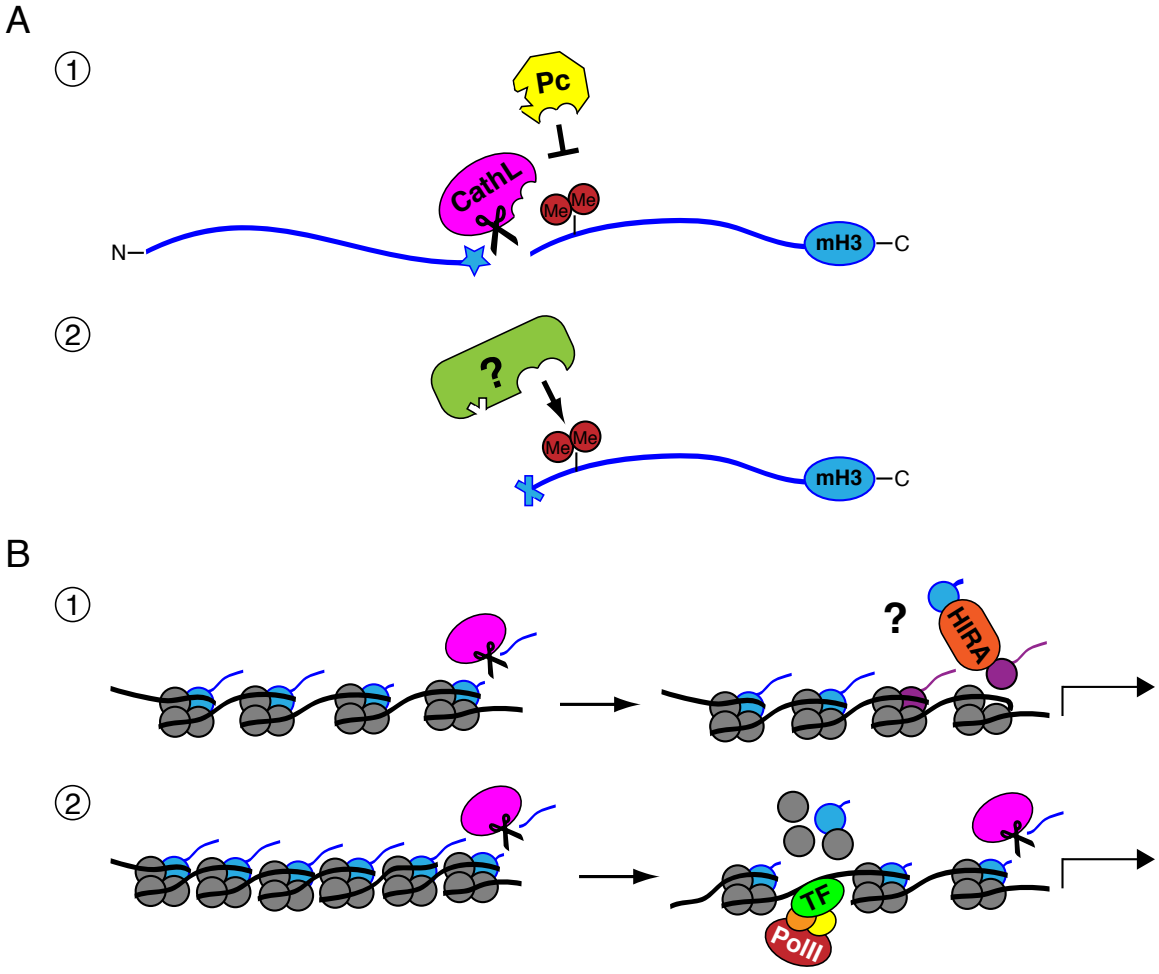


Figure 5.2: Histone cleavage may affect chromatin function through mechanisms operating at the tail and/or nucleosome levels.

A. Cleavage of the H3 tail has been shown to decrease the binding affinity of mammalian Polycomb (Pc) for K27 methylation (1); it may also create a new context for different binding proteins that could not bind prior to cleavage due to steric effects (2).

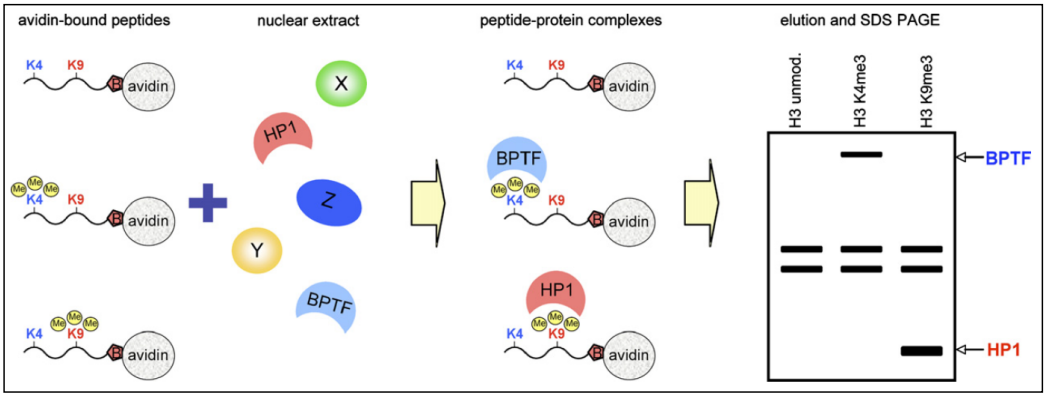
B. H3 cleavage may lead to histone replacement, perhaps mediated by H3.3 chaperone HIRA; such a mechanism might lead to H3.3 accumulation at specific genomic loci (1). Histone proteolysis may also affect nucleosome stability and lead to their eviction; this would then allow greater access of the DNA to transcription factors (TFs), leading to increased transcription at those loci.

demonstrated in Min et al and depicted in **Figure 5.2A-1**, it is also likely that the "new" histone tail produced by its cleavage presents the appropriate context for different chromatin-interacting proteins to bind the same modification. Moreover, the PHD finger of BPTF makes specific contacts with the amino terminus of the H3 tail (as shown in **Figure 2.6C**) leading Ruthenburg et al to hypothesize that the proximity of a given modification to the end of that tail may also affect the ability of particular proteins to bind (e.g. an "end-rule"), in addition to the modification itself and its immediate sequence context (Ruthenburg et al., 2007). In **Figure 5.3** I show some unpublished data suggesting that there are indeed proteins in the nuclei of differentiating ESCs that prefer to bind H3K27me2 and K27me3 in the context of the cleaved peptide (H3 22-37) but not an "uncleaved" peptide (H3 18-37). Using the method established in our laboratory and depicted in **Figure 5.3A**, I performed a screen for such cleaved-histone-methyl-binding proteins and found that many proteins did preferentially associate with the shorter methylated peptides (**Figure 5.3B**). Unfortunately, this experiment did not purify significant quantities of protein for mass spectrometry identification, but it does suggest that such experiments are promising and worth pursuing in future studies.

It is also possible that histone cleavage leads to downstream effects by causing changes at the nucleosome level. Studies have shown that nucleosome composition and positioning on DNA affacts gene expression, but the functional roles of such mechanisms in development and differentiation are not yet fully understood (Rando and Ahmad, 2007). As shown in **Figure 5.2B** and **Figure 1.5A**, cleavage of histone tails (such as that of H3 by Cathepsin L) could initiate other mechanisms of chromatin remodeling, such as histone replacement (**Figure 5.2A-1 and Figure 1.5A**) and/or nucleosome eviction (**Figure 5.2B-2**). Chromatin Immunoprecipitation (ChIP) studies using cleavage-site-specific antibodies and epitope tagged H3 variants, such as those suggested in the Chapter 4 discussion,

could address the possibility that histone cleavage leads to histone replacement at specific genomic loci, as depicted in **Figure 5.2B-1**. In the long term, once the functional roles of histone variant H3.3 in ESC differentiation are better understood, it would be interesting to test whether the perturbation of histone cleavage activity affects some of these downstream functions; such an experiment would address whether there is a functional, mechanistic link or merely a correlative one.

A



from J. Wysocka, Methods, 2006

B

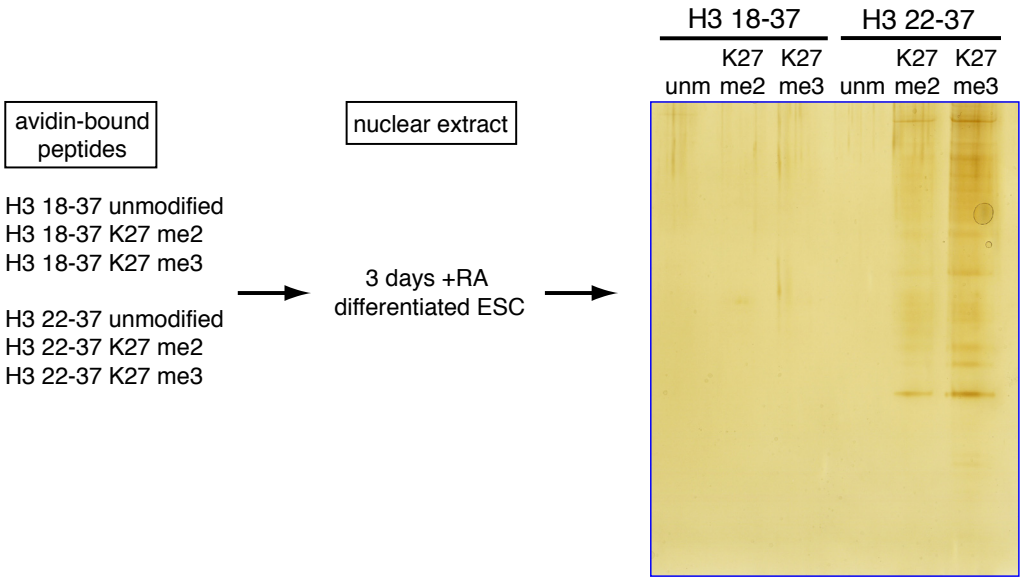
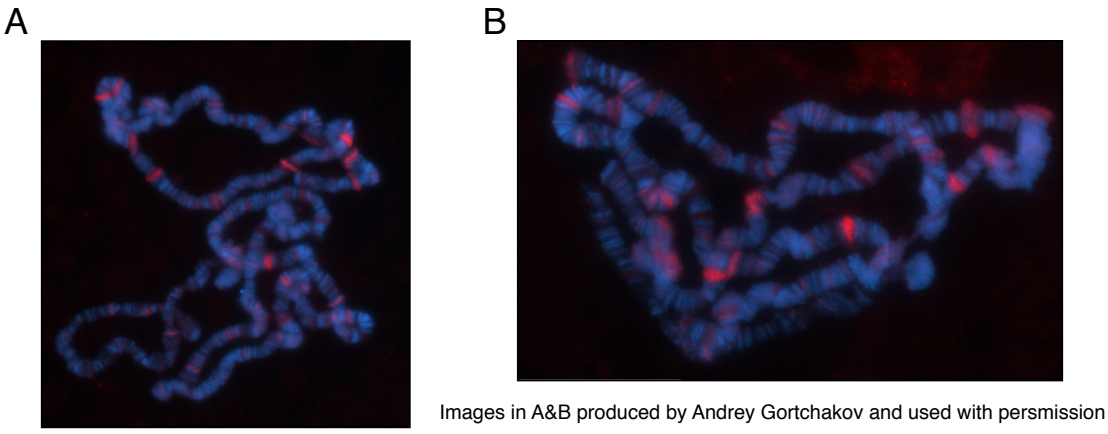


Figure 5.3: H3 peptides mimicking H3 cleavage associate with a unique group of proteins
A. A schematic depicting the peptide pulldown method used to identify novel histone modification binding proteins (Wysocka, 2006).
B. A preliminary experiment showing that cleaved H3 peptides may interact with histone binding proteins that cannot bind to peptides that contain residues amino to the H3 cleavage site.

It may also be possible to test this histone replacement hypothesis in *Drosophila*. Preliminary evidence produced by the Kuroda Laboratory at Harvard Medical School suggests that the cleavage site mapped in differentiating mouse ESCs (**Figure 4.3**) may be conserved in the fly as well. Using the cleavage-specific antibody H3cs.1 in immunofluorescent (IF) labeling of squashed *Drosophila* polytene chromosomes, they observe a pattern that not only suggests the H3cs.1 epitope is present, but also that it is non-randomly distributed (**Figure 5.4 and personal communication with Gortchakov and Kuroda**). Their IF labeling also indicates that only a minor fraction of chromatin is recognized by the H3 cleavage site antibody, which agrees with the data on differentiating ESCs. Since H3 variant replacement was originally documented in *Drosophila* and the fly is a genetically tractable organism, these results suggest that this may be a very useful model in which to test the histone replacement hypothesis described in **Figure 5.2B-1**. Although it is still unclear whether this particular antibody can be optimized for ChIP, the preliminary IF results shown in **Figure 5.4** are suggest that such experiments may be possible.



Images in A&B produced by Andrey Gortchakov and used with persmission

Figure 5.4: The cleavage-specific antibody H3cs.1 labels *Drosophila* polytene chromosomes in a non-random pattern.

A. Immunofluorescent (IF) staining of squashed *Drosophila* polytene chromosomes with the H3 cleavage-specific antibody (α-H3cs.1, shown in red). DNA is stained with DAPI (blue).

B. A second IF image of a chromosome squash (as described in **A**) providing a closer view of the distribution of α-H3cs.1 (red). DNA is stained with DAPI (blue). (Images in **A**&**B** produced by Andrey Gortchakov and used with permission.)

Another interesting model depicting a potential downstream effect of histone cleavage is presented in **Figure 5.2B-2**. Although the mechanistic details are not clear, this model is supported by experiments done in *S. cereviseae* (Santos-Rosa et al., 2009). In their study, Santos-Rosa et al use amino (N) and carboxy (C) terminal epitope tagged H3 proteins to perform ChIP studies, which suggest that H3 tails are cleaved at gene promotors (measured by N-terminal-H3-tag ChIP signal reduction) prior to the depletion of the histone core (measured by a decrease in the C-terminal-H3-tag ChIP signal). They also show that yeast strains expressing a mutant copy of H3 (L20A) do not properly activate genes that normally increase in expression under specific conditions (i.e. stationary phase and sporulation). As shown in **Figure 5.2B-2**, histone cleavage might lead to gene activation by causing nucleosome eviction and therefore allowing the underlying DNA sequence to become more accessible to transcription factors (TFs) and, ultimately, RNA polymerase. Their studies also suggest that the yeast H3 cleavage enzyme discriminates against H3 proteins modified with the activating mark H3K4me3, although it is not clear whether this modification directly inhibits the protease itself or is mediated indirectly through other proteins and/or activities in the yeast nuclear extract.

Although the identity of the yeast H3 cleavage enzyme(s) is still unknown, its discovery will be an important step in studying the downstream effects of histone proteolysis. Since yeast are a genetically tractable system, it will be much easier to test hypothetical models, such as those depicted in **Figure 5.2**, using knockout and mutagenesis strategies that are possible in this model organism.

Unique Histone Modification Patterns in Embryonic Stem Cells

As described in Chapter 3, the chromatin of ESCs bears a unique modification pattern that may be critical to their maintenance of pluripotency and ability to differentiate. At the moment, only one specific ESC pattern of modifications has

been identified: termed “bivalent domains,” this pattern is not only unique in that it correlates with pluripotency but also because it is defined by a combination of marks, H3K4me3 and H3K27me3, that were previously thought to be mutually exclusive (Bernstein et al., 2006). The fact that these modifications are both enriched at genomic loci in ESC chromatin is also significant because they 1) occur in regions that are enriched for genes encoding transcription factors associated with embryonic development and 2) are catalyzed by key developmental proteins, trithorax Group (trxG) and Polycomb Group (PcG) proteins, respectively (depicted in **Figure 5.5A**) (Bernstein et al., 2006). Nevertheless, the question remains as to whether other unique modification patterns exist in ESCs and, if so, whether they have redundant or specialized roles.

Since genes within these bivalent domains are expressed at a very low level, Bernstein et al suggest that the H3K4me3+H3K27me3 bivalent domain maintains genes in a “poised” state. Work by Azuara et al supports this theory by showing that ESCs defective in H3K27 methylation display improper gene expression at these loci (Azuara et al., 2006). One hypothesis might be that ESCs do have additional unique modification patterns, whether on another single tail or in a cross-talk across different histones and that such redundancy would make the mechanism of maintaining “poised” genes more robust. As shown in **Figure 5.5**, such a dual pattern system might then have further implications when histone H3 is cleaved during ESC differentiation. In cleaving the histone tail, a “irreversible” post-translational modification, the cell can both quickly remove modifications (e.g. H3K4me3) and “lock in” the chromatin states dictated by the redundant mark (yellow boxes, **Figure 5.5B**). The function of this mark may be reinforced by the loss of effectors such as Pc and/or by the recruitment of different binding proteins to the new amino terminus of H3. Such a mechanism might only exist during a small window of time during development, such as when an ESC loses its pluripotency,

and allow it to “reset” its epigenetic profile before it proceeds to differentiate toward a specific lineage. It is also possible that this cleavage mechanism is repeated in other instances of stem cell differentiation, such as the differentiation of multipotent skin stem cells (see Chapter 4 discussion), and/or that it occurs on other histone tails, such as H2A.

Many questions still remain as to the function, mechanisms and far-reaching implications of histone proteolysis in both ESCs and other cell types and organisms. Most importantly, these studies have focused exclusively on the “normal” state of cellular growth, differentiation and development. In the future, it will be important

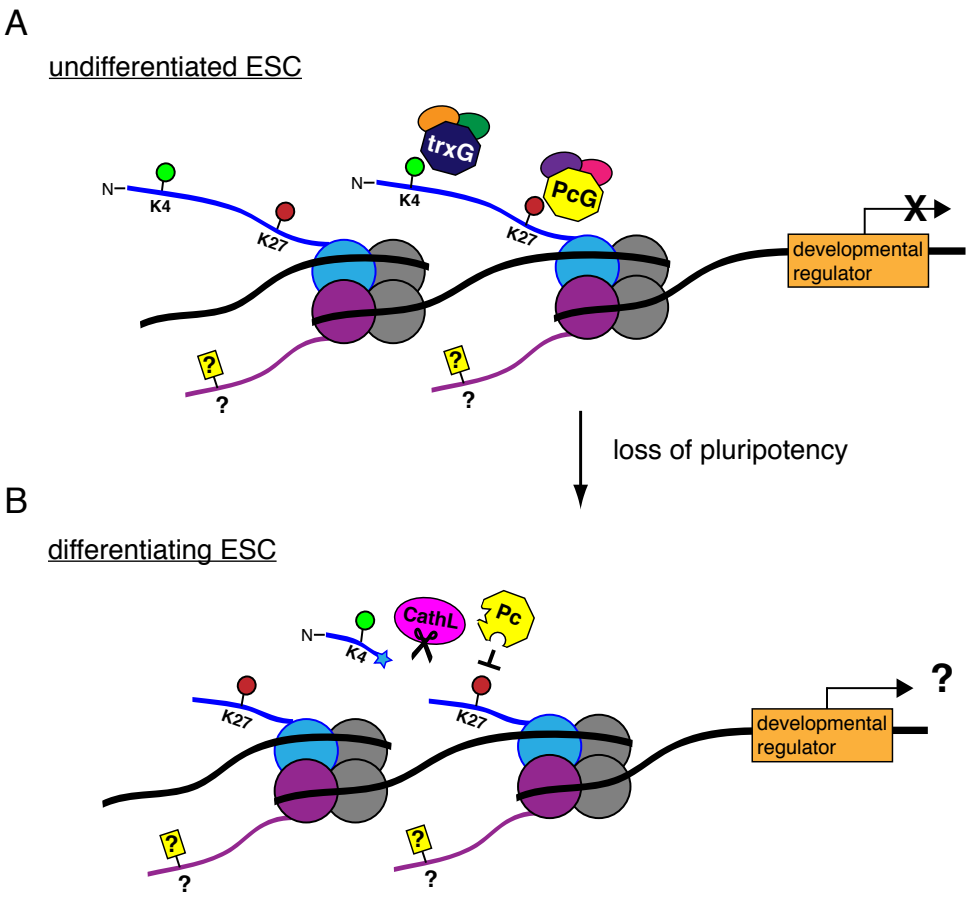


Figure 5.5: The functions of ESC specific modifications may be affected by histone cleavage.

A. A schematic depicting the known ESC-specific modification pattern (H3K4me, represented by the green circle and H3K27me, represented by the red circle) and a potentially unidentified mark. **B.** A schematic showing some of the potential downstream effects on the chromatin status depicted in **A** after H3 tail cleavage.

to address whether this activity exists in disease states and, if so, how it differs from that of normal cells. It is possible that histone cleavage, whether of H3 or other histone tails, could be a clinical marker of certain pathologies and/or lead to testable treatments of such pathologies through the use of enzyme inhibitors. Another important context in which histone cleavage may figure is the recently developed induced Pluripotent Stem cells (iPS), in which differentiated cells are converted to stem cells by the expression of three or four transcription factors (Takahashi and Yamanaka, 2006). The efficiency with which differentiated cells convert to stem-like cells in this process is still quite low, and it is often suggested that this inefficiency is due to the inability of these transcription factors to change the epigenetic state of chromatin. Given that histone cleavage occurs in the process of differentiating pluripotent cells, it may also play a role in the reversal of this process. These questions, and others not raised here, will be the subject of future studies.

Chapter 6: Materials and Methods

Recombinant Protein Preparation

CBX-CDs

CBX CDs (aa 1-62) and full length CBX7 were cloned into pGEX-6P-1 (Amersham); GST fusion proteins were produced in BL21 *E. coli*. Bacterial lysates were purified over Glutathione Sepharose 4B (Amersham) as recommended by the manufacturer. 6x His-tagged human HP1β and *Drosophila* Pc proteins were a gift of W. Fischle (aa 15-72 and aa 15-77, respectively, both cloned into pET-11a; Promega). Point mutations of caging aromatic residues in CBX7 were created using QuikChange site–directed mutagenesis (Stratagene). BPTF PHD Fingers

Wild type and mutant PHD fingers from human BPTF (residues 2583-2640) were expressed and purified as described previously (Li et al., 2006). Briefly, the PHD finger (residues 2583-2640) from human BPTF was cloned into pGex6p vector with an N-terminal glutathione S-transferase (GST) tag. The protein was over-expressed in the *Escherichia coli* host cell Rossetta2 (Novagen) by induction with 0.4 mM isopropyl β-D-thiogalactoside overnight at 20º C in LB medium supplemented with 0.1 mM $ZnCl_2$. After cell lysis and centrifugation, the GST-fusion protein was purified by GST affinity column, followed by tag cleavage using PreScission protease (Amersham Biosciences), and purification by size-exclusion chromatography on Superdex 75. Point mutants of PHD fingers were made using the QuickChange site-directed mutagenesis kit (Stratagene) and were verified by DNA sequencing. All mutants were purified as described for the native protein.

L3MBTL1 MBT repeats

Wild type and mutant 3MBT (residues 197-526 of human L(3)MBTL1, gi: 23396689) were expressed and purified as described previously (Wang et al., 2003).

Histone H3-HIS

A PCR fragment of the complete mouse histone H3.2 ORF was cloned into the pET30a plasmid vector to add a C-terminal HIS tag to the coding sequence. Mutations were made using the Quick-Change Mutagenesis II kit (Stratagene). Plasmids were transformed into BL21 *E.coli* and rH3-HIS protein was purified from inclusion bodies using both Ni^{2+}-NTA and C8 columns (RP-HPLC). Acetic anhydride treatment was performed as described previously (Garcia et al., 2007). rH3-HIS was converted to rH3-HIS+K27me2 by first mutating K27 to C and C110 to A and then alkylating the cysteine to dimethyl by published methods (Simon et al., 2007).

Peptide Synthesis

Peptides were synthesized either at The Rockefeller University Proteomics Resource Center, the Memorial Sloan-Kettering Cancer Center for Microchemistry, or the Baylor College of Medicine Protein Chemistry Facility. The methylation state of each peptide was confirmed by mass spectroscopy. Peptides were extensively lyophilized and resuspended in milli-Q water before use. 5-Carboxyfluorescein (5-FAM) or biotin groups were coupled directly to the N-terminus (for H3 aa 19-35 series) or C-terminus (for H3 aa 1-15 series).

Fluorescence polarization

Fluorescence polarization assays were performed essentially as described (Fischle et al., 2003). All mouse Pc CDs were analyzed using buffer containing 20mM imidazole, 25 mM NaCl, and 2 mM DTT and mHP1α was performed using a buffer containing 50mM Na_2HPO^4, pH=7, 25mM NaCl, 1mM $MgCl_2$, 2mM DTT. The wild-type and mutant PHD fingers were assayed using buffer containing 10mM HEPES pH 7.5, 150 mM NaCl, 0.005% NP-20 and 2 mM DTT. The assays on L3MBTL1

were performed under conditions of 20 mM imidazole, 25 mM NaCl, 2 mM DTT (pH 8.0) or 25 mM Tris-HCl, 2 mM DTT (pH 8.0). 50ml reactions were incubated with 100 nM C-terminally fluorescein-labeled peptide for 15 minutes on ice (Pc/HP1-CDs) or 1 hour at room temperature (PHD and L3MBTL1); fluorescence polarization was then measured in the Hidex Chameleon plate reader, at room temperature.

Fluorescence polarization statistical analysis

For all fluorescence polarization experiments, three independent batches of protein were purified and used in 2-4 replicate binding assays per batch (unless stated otherwise in the text or figure fegend). Anisotropy results for each replicate were normalized and then plotted; binding curves were fit to all normalized data points to determine K_ds and error (SEM). Normalized results for all replicates were averaged and plotted using the Kalidograph software package, with error bars representing standard deviation when included. P value was calculated using the Student's t-test.

Peptide pull-down assays

Peptide pull-down assays were performed as described (Wysocka, 2006) using biotinylated peptides conjugated to streptavidin agarose beads (Pierce 20349). ~5ug of recombinant protein plus 20-fold excess of BSA (internal control) or nuclear extracts from ESCs were incubated with the peptide-conjugated beads in assay buffer (150mM KCl, 20mM Hepes and 0.2% TritonX-100) for 1-4 hours at 4°C; the beads were then washed 6-8x in assay buffer. Elutions were loaded onto 10 or 15% SDS-PAGE gels and silver stained.

Crystallization, data collection and structure determination of BPTF-PHD

Crystallization, data collection and structure determination of the PHD finger of BPTF in complex with H3K4me3 peptide were performed as described previously (Li et al., 2006). Briefly, crystals of the complex between SeMet-labeled PHD finger-bromodomain and $H3_{1-15}$K4me3 were obtained using Crystal Screen Cryo Kit (Hampton Research) condition #41. The protein and peptide were mixed with a ratio of 1:1.2 at ~15 mg/ml and put on ice for ~30 min prior to use. Overnight-grown crystals were briefly cryo-protected by the reservoir solution and frozen under liquid nitrogen. The diffraction data was collected at wavelength 1.3477 Å at the Advanced Photon Source (APS) beam line 23ID-D, Argonne National Laboratory. Due to crystal decay, two data sets from different regions of one crystal were collected with 250x0.5° and 180x1° frames each, and merged for structure determination. All data were processed with HKL2000 (http://www.hkl-xray.com/).

Crystallization, data collection and structure determination of the Y17E PHD finger mutant in complex with H3K4me2 peptide were performed as described previously (Li et al., 2007). Briefly, crystallization was performed by vapor diffusion at 20°C. Y17E sample (16.5mg/ml, 20mM Tris-HCl pH 7.5, 50mM KCl) was pre-incubated with three-fold molar excess of H3(1-9)K4me2 peptide in the presence of 5mM $MgCl_2$ for ~30 min on ice. Crystals were flash-frozen in liquid nitrogen with the reservoir solution as cryoprotectant. Diffraction data of both crystals were collected at beam line NE CAT 24ID-C, the Advanced Photon Source (APS), Argonne National Laboratory, and were processed with HKL2000 (http://www.hkl-xray.com). The phases of both crystals were determined by molecular replacement using the program Molrep associated with the CCP4 software package (http://www.ccp4.ac.uk/html/molrep.html).

Immunoblots

Samples were separated on SDS/PAGE mini gels and transferred onto polyvinylidene difluoride (PVDF) membranes (Millipore); membranes were often then stained with Ponceau S (Sigma) or Amido Black (Sigma) to ensure proper protein transfer. After incubation with primary antibody and addition of a horseradish peroxidase-conjugated secondary antibody (Anti-Rabbit HRP Immunoglobulins, Dako P0399; ECL HRP-Linked Mouse IgG whole Ab, GE NA931V; Polyclonal Rabbit Anti-Goat Immunoglobulins, Dako P0449), membranes were incubated with ECL-Plus substrate (Amersham Pharmacia), and proteins were detected by exposure to x-ray film or using the Fujifilm LAS4000 Imager.

Antibodies

The following antibodies were purchased from commercial vendors: monoclonal α-Flag (M2, Sigma), H3K27me2 (Active Motif 07-452), H3K27me3 (Active Motif 07-449), monoclonal α-Hp1β (Active Motif), monoclonal α-GFP (Roche); Cathepsin L (R&D Systems AF1515), Cathepsin B (R&D Systems AF965), Penta-HIS HRP Conjugate Kit (Qiagen 34460), Oct3/4 (BD Transduction Laboratories 611202), H3gen (Abcam 1791), and H3K4me3 (Abcam 8580). WDR5 was a gift from J. Wysocka and monoclonal H3K27me3 for IF a gift from D. Reinberg. The cleavage-specific H3.cs1 antibody was generated as follows: a 2x branched peptide corresponding to mammalian histone H3 sequence 22-26 was conjugated to KLH and injected into rabbits (Covance). The N-terminal H3gen antibody was generated as follows: a peptide corresponding to mammalian histone H3 sequence 1-6 was conjugated to KLH and injected into rabbits (Covance). Serum was collected and specificity is shown in **Figure 4.6**.

ELISA

Peptides (untreated or those used as substrates in H3 cleavage reactions) were plated onto clear, high-binding ELISA plates in duplicate. Standard curves were generated by plating positive control peptides in duplicate in serial dilutions. Plates were then incubated overnight at 37°C to allow peptides to bind. Each well was then washed 3x with 1X phosphaste buffered saline (PBS). Blocking solution [1% bovine albumin serum (BSA) in 1X PBS + 0.5% Tween-20(PBST)] was then added to each peptide-containing well and the plates were incubated for 1 hour at 37°C. Blocking solution was then removed and replaced with primary antibody diluted in PBST (at a dilution previously optimized for immunoblotting or at several dilutions if optimization is necessary) and incubated for 2-3 hours at 37°C. The wells were then washed 3x with PBST. HorseRadish-Peroxidase (HRP) conjugated secondary antibody was then diluted in 1XPBST, added to each well (1:5000 dilution used for anti-rabbit secondary antibody in ELISAs with H3cs.1 antibody), and incubated for 1-2 hours at 37°C. Wells were then washed 3x with 1XPBST. To develop the assay, OPD tablets (Sigma-Aldrich) were dissolved in MilliQ water and the solution was added to each well and incubated at room temperature for 5-10 minutes (until solution turned yellow in positive control wells). The reactions were then quench by added an equal volume of 1M HCl and absorption measured at 495nm on the Hidex Chameleon plate reader.

Preparation of Histones

Nuclei and histones were isolated as described (Shechter et al., 2007). Cell nuclei were isolated by hypotonic lysis in buffer containing 10 mM Tris-HCl, pH 8.0, 1 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT, 0.4 mM PMSF, protease and phosphatase inhibitors. Pelleted nuclei were acid extracted using 0.4 N sulfuric acid followed by precipitation with trichloroacetic acid or dialysis.

H3 Variant separation by Reverse Phase HPLC (RP-HPLC)

Separation of mammalian core histones by RP-HPLC was done as described (Shechter et al., 2007). Briefly, acid-extracted histones were separated by RP-HPLC on a C8 column (220 by 4.6 mm Aquapore RP-300, Perkin Elmer) using a linear ascending gradient of 35-60% solvent B (solvent A: 5% acetonitrile, 0.1% TFA, solvent B: 90% acetonitrile, 0.1% TFA) over 75 minutes at 1.0 ml/min on a Beckman Coulter System Gold 126 Pump Module and 166/168 Detector. Under these conditions histones H3 split into two peaks. The H3 containing fractions were dried under vacuum and stored at –80°C. RP-HPLC fractions were resuspended in water, analyzed by SDS/PAGE and control-stained with Coomassie brilliant blue. The so identified fractions were then subjected to mass spectrometry (MS) analysis.

Sample Preparation and MS of Histone H3 Variants

Purified histone H3 protein from pooled RP-HPLC fractions were derivatized by treatment with propionyl anhydride reagent (Johnson et al., 2004). The reagent was created using 75 μL of MeOH and 25 μL of propionic anhydride (Aldrich, Milwaukee, WI). Equal volumes of reagent and each H3 variant were mixed and allowed to react at 51°C for 15 min. Propionylated histone H3s were then digested with trypsin (Promega, Madison, WI) at a substrate:enzyme ratio of 20:1 for 5 hours at 37°C after dilution of the sample with 100 mM ammonium bicarbonate buffer solution (pH = 8). The reaction was quenched by the addition of concentrated acetic acid and freezing. A second round of propionylation was then performed to propionylate the newly created peptide N-termini. Propionylated histone digest mixtures were then loaded onto capillary precolumns (360 μm o.d. x 75 μm i.d., Polymicro Technologies, Phoenix, AZ) packed with irregular C18 resin (5-20 μm, YMC Inc., Wilmington, NC) and washed with 0.1% acetic acid for 10 minutes.

Precolumns were connected with Teflon tubing to analytical columns (360 μm o.d. x 50 μm i.d., Polymicro Technologies, Phoenix, AZ) packed with regular C18 resin (5 μm, YMC Inc., Wilmington, NC) structured with an integrated electrospray tip as previously described (Martin et al., 2000). Samples were analyzed by nanoflow HPLC-μ-electrospray ionization on a linear quadrupole ion trap-Fourier Transform Ion Cyclotron Resonance (LTQ-FT-ICR) mass spectrometer (Thermo Electron, San Jose, CA). The gradient used on an Agilent 1100 series HPLC solvent delivery system (Palo Alto, CA) consisted of 0-40% B in 60 min, 45-100% B in 15 min (A = 0.1% acetic acid, B = 70% acetonitrile in 0.1% acetic acid) or other similar gradients. The LTQ-FT mass spectrometer was operated in the data-dependent mode with the 10 most abundant ions being isolated and fragmented in the linear ion trap. All MS/MS spectra were manually interpreted.

2 Dimensional Triton-Acid-Urea (2D TAU) Gels

Total histones were dried under vacuum and resuspended in loading buffer (6 M urea, 0.02% (w/v) Pyronin Y, 5% (v/v) acetic acid, 12.5 mg/ml protamin sulfate). Samples were separated on TAU mini-gels (15% PAGE, 6 M urea, 5% acetic acid, 0.37% Triton X-100; 300 V in 5% acetic acid for 1.5 h). Lanes containing the samples were cut out, adjusted in 0.125 M Tris, pH 6.6 and the TAU gel slice was assembled on top of a 15% SDS/PAGE mini gel. After the run, the gel was stained with Coomassie and destained with 5% methanol, 7.5% acetic acid overnight. The gels were scanned and quantified using Image Gauge software (Science Lab), with the subtraction of background staining.

Histone purification, Edman degradation and MS-MS mapping of H3 cleavage sites

Histones were acid extracted from nuclei and purified by RP-HPLC using a C8 column as described (Shechter et al., 2007). RP-HPLC fractions containing the H3 sub-band were pooled and repurified by RP-HPLC using a C18 column. Equal amounts of fractions 52-55 were pooled, separated by SDS-PAGE, and blotted onto 0.2μm pore size membrane (Millipore ISEQ00010); H3 sub-bands were excised and subjected to Edman degradation as described previously (Strahl et al., 1999) using a Procise 492 cLC Sequencer from Applied Biosystems. Fraction 54 was digested with endoproteinase GluC (Roche Diagnostics, Indianapolis, IN) for 4 hours at 37°C (1:20 wt: wt) and loaded onto a C18 packed capillary column as previously described (Martin et al., 2000). Samples were analyzed by nanoflow HPLC-microelectrospray ionization on a linear quadrupole ion trap-Fourier transform mass spectrometer (LTQ-FT; Thermo Electron) for accurate mass and a Thermo LTQ instrument modified for electron transfer dissociation (ETD) and proton transfer charge reduction (PTR) for adequate tandem mass spectrometry (MS/MS) as previously described (Coon et al., 2005; Syka et al., 2004a; Syka et al., 2004b). Data was manually interpreted, as well as searched against an H3 database using OMSSA (Geer et al., 2004).

***In vitro* H3 cleavage assay**

Extracts were incubated in buffer (10mM HEPES, 10mM KCl, 1.5mM $MgCl_2$, 0.34M sucrose, 10% glycerol, 5mM β-mercaptoethanol, all final) with 0.1μg/μl rH3-HIS (purified from *E. coli*) and incubated at 37°C for 1-3 hours. Reactions were stopped by addition of 5X SDS sample buffer and boiling; results were analyzed by immunoblotting with HIS-HRP and/or H3.cs1 antibody. For MS analysis, reactions were quenched by addition of 0.1% (final) TFA. rCathepsin L (R&D Systems 1515-

CY-010), rCathepsin B (R&D Systems BAF965), and rCathepsin K (Calbiochem 342001) were purchased from commercial sources and tested in the assay with rH3-HIS.

Enzyme enrichment and identification

3 days +RA differentiating ESC chromatin pellets were solubilized by sonication in buffer A (80mM $NaPO^{-4}$, 200mM NaCl), injected onto a 2 or 5mL hydroxyapatite column (BioRad), and fractionated with a 200mM to 2M NaCl gradient. Each fraction was then assayed as described above; both active and inactive fractions were reduced with 1mM DTT at 51°C for 1 hr and alkylated with 2mM iodoacetamide in the dark at RT for 45 mins, followed by digestion with trypsin (Promega Corp., Madison, WI) at an enzyme to substrate ratio of 1:20 (wt:wt) for 6 hrs at 37 °C; digest was then acidified with glacial acetic acid and aliquots of samples were analyzed using an LTQ-FT as described above. Data was searched against a mouse database using SEQUEST (Eng et al., 1994). All spectra of interest were manually validated.

Cellular extract preparation

Whole cell extracts were prepared by resuspending cell pellets in SDS-Laemmli sample buffer and boiling immediately. Chromatin extracts were prepared by sonicating chromatin pellets in buffer (10mM HEPES, 10mM KCl, 1.5mM $MgCl_2$, 0.34M sucrose, 10% glycerol, 5mM β-mercaptoethanol) after isolation by various methods (high salt extraction, (Dignam et al., 1983)**,** chromatin fractionation**, (**Mendez and Stillman, 2000)**)**. Briefly, cells were swelled in low salt buffer and then lysed by either mechanical disruption (Dignam method) or detergent (Stillman method); soluble nuclear proteins were then either extracted with high-salt (Dignam) or released by incubation in no salt buffer and mechanical disruption

(Stillman), leaving behind the chromatin pellet. Protease inhibitors were omitted from extract preparations used for activity assays. Digestion of chromatin fractions into mononucleosomes was accomplished by treatment with microccocal nuclease as described (Wysocka et al., 2001).

Chromatin fractionation and RNAse treatment

Biochemical fractionation was performed by centrifugation as described (Mendez and Stillman, 2000; Wysocka et al., 2005) and depicted in **Figure 2.4A**. Chromatin from female ESCs was prepared at day 6 of differentiation as described above, resuspended in low salt buffer (10mM Hepes, 10mM KCl, and 1.5mM $MgCl_2$), and treated with RNase A at 0, 2, or 5 μg per 100ul of chromatin. Samples were incubated at 37°C for 10 min. and chromatin was run on SDS-PAGE for immunoblots.

Chromatin ImmunoPrecipitation (ChIP)

Chromatin ImmunoPrecipitations were performed essentially as described previously (Lee et al., 2006). Briefly, ESCs were trypsinized, resuspended in warm serum-containing media, and then cross-linked and fixed with 1% paraformaldehyde (final) for 30 minutes at room temperature. The cells were then pelleted and lysed in buffers containing protease inhibitors (Complete Mini tablet, Roche), and sonicated in the Bioruptor (Diagenode) to shear the chromatin. 1/10 volume of 10% Triton X-100 was then added to the sonicated lysates, followed by centrifugation to clear any non-solubilized material. The supernatants were then aliquoted into input and ChIP samples. Pre-antibody-bound Protein A and/or Protein G magnetic beads (Dynabead) were then added to the ChIP samples and incubated overnight at 4°C. The beads were then washed several times and the chromatin fragments eluted. The eluates were then treated to reverse the protein-DNA cross-linking and digest the remaining protein and RNA. The remaining genomic DNA was then purified

using the Qiagen PCR purification kit and split into aliquots for QPCR and Solexa sequencing. Those aliquots to be sent for sequencing were prepared by first using the Epicentre DNA ENDRepair kit (Epicentre Biotechnologies, Cat. # ER0720) to generate blunt-ended DNA, then treated with the Klenow enzyme (New England Biolabs) to add adenines to the 3' ends, ligated to the Solexa adapter using T4 ligase (New England Biolabs), and amplified by PCR. Amplified DNA was then sequenced using the Ilumina Genome Analyzer (Solexa) in the Genomics Resource Center at The Rockefeller University.

RNA gel shifts

RNA was *in vitro* transcribed with T7 Megascript (Ambion) and ^{32}P-UTP. 100 and 500 nt fragments of cyclin E were used as a template for single and double-stranded RNAs. RNAs were gel purified. Cyclin E DNA was labeled with ^{32}P γ-ATP using T4 Polynucleotide kinase. Increasing amounts of protein (10-500 pmol) were incubated in buffer containing 20mM Hepes, 100mM KCl, 2mM EDTA, 0.01% NP-40, 1mM DTT, 3ug tRNA for RNA gel shifts (Ambion), 1ug poly (dI-dC) for DNA gel shifts (Amersham), RNasin (Promega) and 10,000 cpm of RNA or DNA in a total reaction volume of 20 μl. Reactions were incubated on ice for 30 min.; 5% native 0.5xTBE gel was pre-run at RT; gels were run at 4°C for 3 hours at 250v and imaged with FLA-5000 phosphorimager (Fujifilm).

Cell Lines and Culture

All mammalian cell lines, with the exception of mouse LF2 cells were grown in Iscove's DMEM supplemented with 10% fetal calf serum and penicillin/streptomycin at 37^{o}C and 5% CO_2. The following cell lines were used in this study: HeLa (human cervical epithelium, gift from Lisa K. Denzin, MSKCC), HT-29 (human colon cancer, gift from Petr Protiva, Cornell University), Raji (human Burkitt's lymphoma,

gift from Lisa K. Denzin, MSKCC), CEM (human T cell leukemia lymphocyte, gift from Lisa K. Denzin, MSKCC), HEK293 (human embryonic kidney), U2OS (human osteosarcoma, ATCC#: HTB-96), MEF (mouse embryonic fibroblast, gift from Andre Nussenzweig, NCI), LF2 (mouse embryonic stem cell, gift from Austin Smith, CSCR, UK), N2a (mouse neuronal stem cell-like, gift from Robert Darnell, Rockefeller University), P-CUT MEF and 10T1/2 (mouse embryonic fibroblast).

ESCs (LF2 line, see above) were cultured as previously described (Bernstein et al., 2006; Smith, 1991). Cells were grown on gelatin-coated plates without feeder cells and maintained in an undifferentiated state through culture in KODMEM (Invitrogen 1082-9018), 2mM L-glutamine (Sigma G7513), 15% ES grade fetal bovine serum (Gibco 10439-024), 10^{-4} mM 2-mercaptoethanol, and leukemia inhibitory factor (LIF). Cells were plated and maintained at 60-80% density, requiring passage every 2-3 days and checked for the expression of positive stem cell markers by immunoblot periodically. To differentiate, cells were plated at ~$1*10^4$ cells/cm^2 and differentiated with LIF-free media ± 100nM all-trans-RA (Sigma R2625). Differentiation ES medium was changed every day for the first three days, during which time RA was maintained. After three days, RA was removed and cells cultured with differentiation ES medium that did not contain RA. Embryoid bodies were formed by splitting partially trypsinized cells onto non-treated petri dishes and allowing cells to cluster. Cathepsin L was chemically inhibited by adding 10uM [final] Cathepsin L I inhibitor (Calbiochem 219402) to the cell media.

CBX cDNA fusions

CBX cDNAs were cloned BglII/SalI into EGFP-N1 (BD Biosciences) producing C-terminal EGFP fusion proteins. CBX2 cDNA was a gift of M. Narita. CBX4 cDNA was cloned from female ES cell line HP310 by RT-PCR; cDNAs for CBX6 and

CBX8 were purchased from Open Biosystems; CBX7 cDNA previously described (Gil et al., 2004). All CBX cDNAs were verified. Point mutations in CBX7-EGFP were created as for recombinant proteins. The CBX4 CD swap into CBX7 was made by staggered PCR by incorporation of overlapping oligos and the chimeric cDNA was cloned into EGFP-N1.

Immunofluorescence and RNase treatment

LF2 cells were transiently transfected at day 2 of differentiation with Lipofectamine 2000 (Invitrogen) and on day 3, cells were stained as previously described (Chaumeil et al., 2002). Line histograms were made with processing software, MetaMorph Offline. RNase treatments on living LF2 cells were performed as described (Muchardt et al., 2002). In brief, 24 hours post-transfection, cells were incubated on ice for 10 minutes in PNS-0.3% Triton X-100 (PNS=20 mM PIPES pH 6.8, 200 mM NaCl, 600 mM sucrose). Cells were incubated for 1 hour on ice in PNS +/- RNase A (1 mg/ml) and fixed in 4% paraformaldehyde. IF was performed as above.

Cell Cycle Analysis by FACS

$1x10^6$ cells were collected, washed with PBS and fixed overnight at -20ºC in 70% Ethanol, diluted in PBS. The next day, cells were washed with PBS and incubated for 30 min at 37ºC in PBS containing RNaseA (10μg/ml), followed by the addition of propidium iodide (10μg/ml) and another incubation for 30 min at 37ºC. Stained cells were analyzed on a FACSort instrument (BD Immunocytometry Systems) with the exclusion of doublets. Analysis of the results was performed with CellQuest software (BD Bioscience).

RNAi knockdown

Short hairpins to both a control gene (a human gene that does not share sequence homology to the mouse genome) and the *Ctsl* gene were purchased from Open Biosystems (RHS3979-9628371 and RMM3981-9597987) and transfected into 293T cells to produce Lentiviral particles. ESC lines were created by infection and selection with puromycin. Heterogeneous cell populations expressing either the control or Ctsl RNAi were then differentiated with RA.

RNA isolation, reverse transcription and Q-PCR with SYBRgreen

RNA was isolated from cells using Trizol reagent (Invitrogen) following the recommended protocol. RNA quality was assessed using the Nanodrop Spectrophotometer (Thermo Scientific) by measuring the 260nm wavelength absorption to ensure that it produced a regular peak, and the 260/280nm ratio to the to ensure that it was > 1.8. RNA that met these requirements was then reverse transcribed into cDNA using oligo dT primer and the SuperScript III First-Strand Synthesis Kit (Invitrogen). Reverse transcribed cDNA was then used in quantitative PCR (Q-PCR) reactions with SYBRgreen reagent and the incorporated fluorescence was measured by the Stratagene Mx3000P QPCR System. Standard curves were generated for each primer pair using serial dilutions of cDNA, and each standard and sample was measured in duplicate. The sequences of all primers used to amplify specific genes are listed in the table below.

Table 1. Sequences of primers used in Q-PCR with SYBRgreen

Gene	Forward seq (5'-3')	Reverse seq (5'-3')	Ann Tmp C
Nanog	CCTTCCCTCGCCATCACACT	AGAGGAAGGGCGAGGAGAGG	57
Nestin	TCGGGAGAGTCGCTTAGAG	AGTTGCTGCCCACCTTCC	57
Pax3	CCATCGGCGGCAGCAAACC	GATCCGCCTCCTCCTCTTCTCCTT	57
Musashi	GCGGCGGCGGCAGCTGTAGTTCG	CGGGGCTGGCGGCGCTGATGTAA	59
HoxA1	AGCCACCAAGAAGCCTGTCGTT	TCTGTGAGCTGCTTGGTGGTGAAA	57
Sox17	CGCGCTCGACGGCTACC	GCGGGCGGCTCTACGGACACT	59
GATA6	TGCCTCGACCACTTGCTATGAAAA	CACTGATGCCCCTACCCCTGAG	55

Appendix: genes marked differentially by H3K4me3

refID = refSeq gene ID; **chr** = chromosom; **strd** = DNA strand; **ES_exp** = ESC sample expression; **und** = undifferentiated ESC sample; **3RA** = 3 days +RA differentiating cell sample; **0** = non-significant sequence representation; **1** = significant sequence representation; **name** = gene name

genes enriched in H3K4me3 in und ESCs (und = 1) are listed first followed by those enriched in H3K4me3 in 3RA ESCs (3RA = 1); within each sub-list, genes are listed in ascending alphabetical order according to gene name

Table 2. Genes marked differentially by H3K4me3 - (pages 132-154)

refID	chr	strd	ES_exp	und	3RA	name
NM_026690	chr7	+	3.819966333	1	0	0610012D14Rik
NM_026886	chr5	-	7.750643333	1	0	1500001A10Rik
NM_028156	chr11	-	9.438215	1	0	1700001P01Rik
NM_030141	chr17	+	35.90888333	1	0	1700061G19Rik
NM_183097	chr8	+	53.16986667	1	0	1700067K01Rik
NM_053216	chr9	+	17.0918	1	0	1700102P08Rik
NM_029607	chr2	+	339.7533333	1	0	2310003C23Rik
NM_029747	chr17	-	28.9215	1	0	2410137M14Rik
NM_175381	chr19	+	51.4109	1	0	2700081O15Rik
NM_001033547	chr1	-	6.88332	1	0	4922505E12Rik
NM_001014836	chr10	-	6.98401	1	0	4930404N11Rik
NM_026296	chr5	+	13.41806333	1	0	4930548H24Rik
NM_027633	chr6	-	2.120500667	1	0	4931417G12Rik
NM_027519	chr5	+	6.018168333	1	0	6330406I15Rik
NM_177265	chr7	-	2.066853333	1	0	6330512M04Rik
NM_183152	chr10	+	53.06503333	1	0	6330514A18Rik
NM_021477	chr16	+	3.638213333	1	0	A2bp1
NM_173008	chr7	+	4.681863333	1	0	A430110N23Rik
NM_178656	chr11	+	10.3785625	1	0	A530088H08Rik
NM_001013793	chr17	+	23.8078	1	0	AA388235
NM_011920	chr6	+	697.8356667	1	0	Abcg2
NM_001033163	chr17	-	20.61596667	1	0	Abhd9
NM_198018	chr11	-	74.74003333	1	0	Abr
NM_134247	chr12	+	3.3122785	1	0	Acot4
NM_145444	chr12	+	3.399465333	1	0	Acot5
NM_172845	chr1	+	18.04867	1	0	Adamts4
NM_197985	chr6	-	389.27725	1	0	Adipor2
NM_009646	chr10	-	184.5286667	1	0	Aire
NM_001029876	chr8	+	387.8326667	1	0	AK122209

refID	chr	strd	ES_exp	und	3RA	name
NM_028270	chr4	+	107.8928	1	0	Aldh1b1
NM_145684	chr11	-	19.08561667	1	0	Alox12e
NM_007441	chr3	+	7.171318833	1	0	Alx3
NM_019871	chr11	-	17.45036667	1	0	Amac1
NM_001012450	chr4	-	60.79077778	1	0	Ankrd6
NM_009676	chr1	+	8.509	1	0	Aox1
NM_031159	chr6	-	78.0922	1	0	Apobec1
NM_001005508	chr1	+	20.61983889	1	0	Arhgap30
NM_001002897	chr5	-	16.90213333	1	0	Atg9b
NM_026094	chr10	-	27.743845	1	0	Atp8b3
NM_001080935	chr11	+	39.98044233	1	0	B230217C12Rik
NM_001033308	chr4	-	28.25383333	1	0	BC013712
NM_144869	chr19	-	61.15306667	1	0	BC021614
NM_153513	chr3	-	49.95230178	1	0	BC028528
NM_009747	chr12	+	2.353836667	1	0	Bdkrb2
NM_001033539	chr16	+	1.567877	1	0	Bex6
NM_010800	chr5	+	17.69241333	1	0	Bhlhb8
NM_175235	chr9	+	7.282936667	1	0	Brunol6
NM_001017525	chr10	+	98.18878889	1	0	Btbd11
NM_001039231	chr5	-	22.27236	1	0	C230055K05Rik
NM_175427	chr2	-	6.553708333	1	0	C630035N08Rik
NM_013879	chr5	-	40.4871	1	0	Cabp1
NM_007583	chr15	-	8.57833	1	0	Cacng2
NM_133746	chr19	-	21.05016667	1	0	Calhm2
NM_011797	chr3	-	12.25247667	1	0	Car14
NM_001033243	chr7	+	12.38032667	1	0	Ccdc114
NM_153784	chr17	+	9.480913333	1	0	Ccdc64b
NM_177600	chr2	+	1.483516667	1	0	Ccdc73
NM_009913	chr9	+	3.231558222	1	0	Ccr9
NM_007645	chr7	-	10.2986	1	0	Cd37
NM_007646	chr5	+	73.28866667	1	0	Cd38
NM_009853	chr11	-	236.2413333	1	0	Cd68
NM_133974	chr9	-	69.13226667	1	0	Cdcp1
NM_019707	chr8	+	7.985309583	1	0	Cdh13
NM_133709	chr7	+	10.03623333	1	0	Chrdl2
NM_145129	chr9	-	4.122701833	1	0	Chrna3
NM_001081104	chr5	+	20.07288167	1	0	Chrna9
NM_019500	chr16	-	14.66071667	1	0	Cldn14
NM_025809	chr12	-	2.062456	1	0	Clec14a
NM_177129	chr1	-	11.78899917	1	0	Cntn2

refID	chr	strd	ES_exp	und	3RA	name
NM_007742	chr11	+	32.236045	1	0	Col1a1
NM_031163	chr15	-	24.47193333	1	0	Col2a1
NM_009946	chr13	+	22.50726111	1	0	Cplx2
NM_170673	chr6	+	57.92613333	1	0	Cpne9
NM_026682	chr11	-	103.5890167	1	0	Cpsf4l
NM_028798	chr3	-	15.19594333	1	0	Crct1
NM_172122	chr4	-	37.48393333	1	0	Crocc
NM_007795	chr7	+	52.33413333	1	0	Ctf1
NM_010009	chr10	+	6.15731	1	0	Cyp27b1
NM_145548	chr4	-	8.25364	1	0	Cyp2j13
NM_028979	chr4	-	12.4709	1	0	Cyp2j9
NM_178727	chr4	-	10.11254667	1	0	D630039A03Rik
NM_007584	chr17	-	66.27872671	1	0	Ddr1
NM_010043	chr1	+	51.33761333	1	0	Des
NM_010056	chr6	-	1.538752667	1	0	Dlx5
NM_001081330	chr11	-	10.69015	1	0	Dnahc2
NM_019874	chr4	+	52.07904444	1	0	Dnajb5
NM_153743	chr12	+	135.4267439	1	0	Dnmt3a
NM_001081695	chr10	+	751.661	1	0	Dnmt3l
NM_010077	chr9	+	4.800603333	1	0	Drd2
NM_025869	chr8	+	11.93795667	1	0	Dusp26
NM_001033344	chr1	-	63.86603333	1	0	Dusp27
NM_007904	chr14	-	8.420666667	1	0	Ednrb
NM_007911	chr11	-	33.2705	1	0	Efnb3
NM_019397	chrX	-	9.15442	1	0	Egfl6
NM_025499	chr10	+	5.099154667	1	0	Eid3
NM_207685	chr4	-	429.9138667	1	0	Elavl2
NM_013514	chr14	-	109.5730167	1	0	Epb4.9
NM_144848	chr15	-	123.0061333	1	0	Eppk1
NM_019585	chr4	-	15.42566833	1	0	Espn
NM_176846	chr9	+	1.076221333	1	0	Exph5
NM_010164	chr1	-	13.02519	1	0	Eya1
NM_007977	chrX	-	11.03175333	1	0	F8
NM_015798	chr18	+	2274.693333	1	0	Fbxo15
NM_001077189	chr1	-	20.10631807	1	0	Fcgr2b
NM_033522	chr12	+	5.125443	1	0	Ferd3l
NM_010194	chr7	-	13.76045	1	0	Fes
NM_080433	chr14	-	5.802793333	1	0	Fezf2
NM_010199	chr16	-	4.5063785	1	0	Fgf12
NM_020013	chr7	-	4.966183333	1	0	Fgf21

refID	chr	strd	ES_exp	und	3RA	name
NM_010208	chr4	+	20.81738333	1	0	Fgr
NM_033571	chr5	-	41.32363333	1	0	Fkbp6
NM_008051	chr7	+	1.031342	1	0	Fut1
NM_008082	chr18	-	4.590435444	1	0	Galr1
NM_008105	chr13	+	111.950565	1	0	Gcnt2
NM_144891	chr2	+	8.3071505	1	0	Gdap1l1
NM_008110	chr11	+	25.21456667	1	0	Gdf9
NM_008120	chr4	-	17.36186667	1	0	Gja4
NM_008126	chr4	-	427.013	1	0	Gjb3
NM_183427	chrX	-	3.7543115	1	0	Glra2
NM_001033452	chr2	-	37.79986667	1	0	Gm1967
NM_010317	chr13	+	6.872251111	1	0	Gng4
NM_021610	chr1	+	90.2146	1	0	Gpa33
NM_146250	chr1	-	17.50173333	1	0	Gpr1
NM_010951	chrX	+	3.08802	1	0	Gpr143
NM_007412	chr10	-	19.67506667	1	0	Gpr182
NM_173410	chr7	+	5.955345	1	0	Gpr26
NM_011825	chr1	-	24.9065	1	0	Grem2
NM_026960	chr15	+	38.3199	1	0	Gsdmd
NM_029555	chr6	+	70.83673333	1	0	Gstk1
NM_026619	chr19	+	15.7455	1	0	Gsto2
NM_008185	chr10	-	55.99386667	1	0	Gstt1
NM_010398	chr17	-	19.94946667	1	0	H2-T23
NM_013548	chr13	+	214.3566667	1	0	Hist1h3f
NM_172563	chr11	-	8.236425	1	0	Hlf
NM_027521	chr10	+	40.827	1	0	Hmha1
NM_175606	chr5	-	62.11116444	1	0	Hopx
NM_026489	chr3	+	121.3566667	1	0	Hormad1
NM_010475	chr11	+	24.44636667	1	0	Hsd17b1
NM_025330	chr7	+	90.38643333	1	0	Hsd17b14
NM_001040684	chr7	+	156.185	1	0	Hsd3b7
NM_008310	chr16	-	2.827358833	1	0	Htr1f
NM_027320	chr11	+	65.4837	1	0	Ifi35
NM_026820	chr7	+	3796.616667	1	0	Ifitm1
NM_025378	chr7	-	3844.43	1	0	Ifitm3
NM_008340	chr17	+	22.32913333	1	0	Igfals
NM_008344	chr15	+	15.3358	1	0	Igfbp6
NM_198610	chr4	-	91.6409	1	0	Igsf21
NM_011772	chr10	-	11.949744	1	0	Ikzf4
NM_016671	chr8	-	4.792233333	1	0	Il27ra

refID	chr	strd	ES_exp	und	3RA	name
NM_134109	chr16	+	74.39956667	1	0	lldr1
NM_001033354	chr6	-	28.59006667	1	0	lqsec3
NM_010592	chr8	+	172.21201	1	0	Jund
NM_021275	chr2	+	6.861653333	1	0	Kcna4
NM_032397	chr8	-	19.99056667	1	0	Kcnn1
NM_001081403	chr18	-	0.993792667	1	0	Klhl14
NM_175174	chr5	+	25.53080844	1	0	Klhl5
NM_001039042	chr7	+	10.86359	1	0	Klk13
NM_010663	chr11	-	14.08779667	1	0	Krt17
NM_027221	chr5	+	109.9396667	1	0	Krtcap3
NM_008479	chr6	-	13.45605667	1	0	Lag3
NM_027570	chr8	-	12.84734833	1	0	Ldhd
NM_144862	chr18	+	29.6078	1	0	Lims2
NM_001031772	chr10	-	29.90056667	1	0	Lin28b
NM_175271	chrX	+	16.45565778	1	0	Lpar4
NM_008511	chr6	+	19.91353333	1	0	Lrmp
NM_027941	chr3	-	1307.973333	1	0	Lrrc34
NM_028175	chr8	+	7.151696667	1	0	Lrrc8e
NM_010734	chr17	-	17.76182	1	0	Lst1
NM_027366	chr17	+	372.157	1	0	Ly6g6e
NM_008535	chr8	+	13.68592	1	0	Lyl1
NM_001080390	chr19	-	77.78282222	1	0	Mark2
NM_001045484	chr8	+	44.2102	1	0	Mef2b
NM_008584	chr12	+	2.677003667	1	0	Meox2
NM_020021	chr4	-	1.910684	1	0	Mos
NM_001081361	chr1	-	11.53505167	1	0	Mosc1
NM_175534	chr7	-	16.04576667	1	0	Mrgpre
NM_018857	chr17	-	29.8378	1	0	Msln
NM_028069	chr7	-	12.30094667	1	0	Mupcdh
NM_146189	chr7	-	18.81796667	1	0	Mybpc2
NM_016754	chr7	+	735.4426667	1	0	Mylpf
NM_028016	chr6	+	2832.333333	1	0	Nanog
NM_201355	chr7	+	15.00292	1	0	Nat14
NM_007789	chr8	-	29.49239	1	0	Ncan
NM_010877	chr1	+	38.15966667	1	0	Ncf2
NM_013609	chr3	+	11.99053	1	0	Ngf
NM_008703	chr10	+	1.38555	1	0	Nmbr
NM_019401	chr2	-	30.72056667	1	0	Nmi
NM_010929	chr17	+	49.09722556	1	0	Notch4
NM_001029877	chr7	+	28.84873833	1	0	Nova2

refID	chr	strd	ES_exp	und	3RA	name
NM_013839	chr2	-	10.10094667	1	0	Nr1h3
NM_010947	chr17	-	1.794152667	1	0	Ntn2l
NM_019738	chr7	-	310.5866667	1	0	Nupr1
NM_001024141	chrX	-	2.191649333	1	0	Nxf3
NM_011025	chr2	+	63.6228	1	0	Oxt
NM_001038839	chr5	+	14.06871889	1	0	P2rx7
NM_011862	chr15	-	802.9116667	1	0	Pacsin2
NM_001033254	chr2	+	9.622628333	1	0	Pak6
NM_019943	chr5	-	9.172696667	1	0	Papolb
NM_011040	chr2	-	9.763525	1	0	Pax8
NM_001003671	chr18	+	4.43982075	1	0	Pcdhac1
NM_008788	chr5	-	670.69	1	0	Pcolce
NM_013628	chr13	+	13.45308333	1	0	Pcsk1
NM_211138	chrX	+	57.02986667	1	0	Pcyt1b
NM_145978	chr14	-	56.55536667	1	0	Pdlim2
NM_018863	chr2	-	6.509916667	1	0	Pdyn
NM_021453	chr19	-	180.5266667	1	0	Pga5
NM_001042623	chr6	-	2727.77	1	0	Phc1
NM_008952	chr11	-	415.5423333	1	0	Pipox
NM_011987	chr16	-	307.1746667	1	0	Pla2g10
NM_011107	chr5	+	236.5109293	1	0	Pla2g1b
NM_177845	chr2	-	14.1262	1	0	Pla2g4e
NM_019588	chr19	+	21.96623333	1	0	Plce1
NM_177355	chr15	+	5.588176667	1	0	Plcxd3
NM_201394	chr15	-	113.3467208	1	0	Plec1
NM_001033253	chr10	-	35.55724667	1	0	Plekhg1
NM_001004156	chr4	+	95.6651	1	0	Plekhg5
NM_153169	chrX	+	2.917573333	1	0	Pnma3
NM_001081209	chr1	-	285.2516667	1	0	Prdm14
NM_027230	chr2	-	123.2719533	1	0	Prkcbp1
NM_001039079	chr4	-	85.09287111	1	0	Prkcz
NM_001081374	chr7	-	122.882	1	0	Prss36
NM_177611	chr2	+	12.58479167	1	0	Psd4
NM_008969	chr2	+	5.123938333	1	0	Ptgs1
NM_008973	chr6	-	7.581533167	1	0	Ptn
NM_013545	chr6	-	137.954	1	0	Ptpn6
NM_023852	chr13	-	3.617692222	1	0	Rab3c
NM_028226	chr4	+	15.836735	1	0	Rbm12b
NM_009049	chr1	-	0.683699333	1	0	Resp18
NM_028713	chr1	-	2.495623222	1	0	Rftn2

refID	chr	strd	ES_exp	und	3RA	name
NM_183163	chr4	+	9.43289	1	0	Rhbdl2
NM_020599	chr7	-	34.8106	1	0	Rlbp1
NM_026301	chr18	+	269.3866667	1	0	Rnf125
NM_199057	chr4	+	73.14975	1	0	Rusc2
NM_015772	chr14	-	25.45373333	1	0	Sall2
NM_009130	chr9	-	3.432306667	1	0	Scg3
NM_146013	chr11	+	32.31553333	1	0	Sec14l4
NM_148942	chr13	-	119.1670667	1	0	Serpinb6c
NM_019982	chr5	-	11.730275	1	0	Sez6l
NM_018780	chr19	-	10.30011	1	0	Sfrp5
NM_177816	chr14	-	5.34699	1	0	Sh2d4b
NM_001081028	chr7	-	71.8359	1	0	Sipa1l3
NM_020333	chr2	+	19.59880556	1	0	Slc12a5
NM_001083902	chr16	+	25.11923	1	0	Slc12a8
NM_022411	chr11	-	9.612196667	1	0	Slc13a2
NM_053079	chr14	-	41.56753333	1	0	Slc15a1
NM_080853	chr7	+	2.828433333	1	0	Slc17a6
NM_173774	chr4	-	41.82163333	1	0	Slc45a1
NM_053077	chr15	+	10.73790667	1	0	Slc45a2
NM_139142	chr9	-	2.142096667	1	0	Slc6a20a
NM_172861	chr3	-	3.139431667	1	0	Slc7a14
NM_199065	chr14	-	6.191556667	1	0	Slitrk1
NM_146064	chr15	+	23.5759	1	0	Soat2
NM_009235	chr11	+	180.655	1	0	Sox15
NM_001025559	chr7	-	5.789103833	1	0	Sox6
NM_009263	chr5	+	4039.823333	1	0	Spp1
NM_053188	chr17	-	16.15736667	1	0	Srd5a2
NM_016911	chrX	-	8.859126667	1	0	Srpx
NM_175162	chr8	-	40.7198825	1	0	Stox2
NM_172294	chr1	+	52.34398583	1	0	Sulf1
NM_027890	chr10	-	135.4453333	1	0	Susd2
NM_153579	chr7	-	7.300501667	1	0	Sv2b
NM_026805	chr5	-	9.40786	1	0	Svop
NM_011523	chr17	+	16.26727517	1	0	Synj2
NM_009308	chr18	-	12.05706833	1	0	Syt4
NM_016908	chr7	-	14.34537667	1	0	Syt5
NM_011536	chr11	+	11.55927833	1	0	Tbx4
NM_013836	chr15	-	89.70518889	1	0	Tcf20
NM_009337	chr12	-	1014.861333	1	0	Tcl1
NM_013688	chr17	+	23.71636667	1	0	Tcte1

refID	chr	strd	ES_exp	und	3RA	name
NM_175030	chr4	+	9.883396667	1	0	Tctex1d4
NM_011562	chr9	-	6064.253333	1	0	Tdgf1
NM_013690	chr4	+	25.03686667	1	0	Tek
NM_011569	chr11	-	25.448	1	0	Tekt1
NM_009384	chr16	-	65.58181667	1	0	Tiam1
NM_001037744	chr14	+	7.006266667	1	0	Timm8a2
NM_181856	chr11	+	45.72756667	1	0	Tmc8
NM_175408	chr6	+	15.03183333	1	0	Tmem139
NM_178381	chr7	-	147.8876667	1	0	Tmem16j
NM_178915	chr12	-	17.69183333	1	0	Tmem179
NM_144936	chr9	-	14.16273333	1	0	Tmem45b
NM_025452	chr4	+	20.3242	1	0	Tmem54
NM_001033334	chr12	-	8.828598167	1	0	Tmem90a
NM_011609	chr6	+	78.80666667	1	0	Tnfrsf1a
NM_033042	chr4	+	7.575616667	1	0	Tnfrsf25
NM_009401	chr4	-	15.6071	1	0	Tnfrsf8
NM_023517	chr11	-	151.102	1	0	Tnfsf13
NM_033622	chr8	+	8.671323444	1	0	Tnfsf13b
NM_031880	chr11	-	49.49483333	1	0	Tnk1
NM_001025261	chr3	-	963.531775	1	0	Tpd52
NM_009426	chr6	-	375.4086667	1	0	Trh
NM_181853	chr7	-	2.693260667	1	0	Trim66
NM_053166	chr11	+	7.347169	1	0	Trim7
NM_001035239	chr19	+	3.242314167	1	0	Trpm3
NM_022017	chr5	-	3.71322	1	0	Trpv4
NM_028417	chr7	+	9.24646	1	0	Ttc9b
NM_172818	chr15	-	32.04264	1	0	Ttll8
NM_134028	chr11	+	7.185812667	1	0	Tubg2
NM_001009573	chr11	-	25.4736	1	0	Unc13d
NM_152823	chr17	+	19.22723333	1	0	Unc5cl
NM_021397	chr7	-	56.149145	1	0	Zbtb32
NM_009540	chr10	-	34.00788889	1	0	Zfa
NM_022985	chr7	-	720.0477246	1	0	Zfand6
NM_028913	chr7	+	139.16	1	0	Zfp819
NM_009567	chr7	+	33.89545	1	0	Zfp93
NM_001039242	chr11	-	9.701914167	1	0	
NM_017461	chr7	-	66.57913333	1	0	
NM_029687	chr2	+	18.79973167	1	0	
NM_026152	chr19	+	17.11593333	0	1	0610010D20Rik
NM_028747	chr2	+	5.25449	0	1	0610012H03Rik

refID	chr	strd	ES_exp	und	3RA	name
NM_001081123	chrX	+	4.789186667	0	1	1100001E04Rik
NM_025427	chr14	-	112.2610333	0	1	1190002H23Rik
NM_029639	chr9	+	54.8588	0	1	1600029D21Rik
NM_001024614	chr5	+	4.915230333	0	1	1700007G11Rik
NM_025851	chr17	+	4.274049444	0	1	1700010I14Rik
NM_001017407	chr10	+	4.903274833	0	1	1700021K02Rik
NM_029338	chr17	-	2.929013333	0	1	1700027N10Rik
NM_026931	chr8	-	21.886605	0	1	1810011O10Rik
NM_025459	chr15	+	95.92802189	0	1	1810015C04Rik
NM_027363	chr7	+	6.496243333	0	1	2010110P09Rik
NM_001081226	chr1	-	44.64756667	0	1	2310009B15Rik
NM_001077348	chr17	-	22.78533333	0	1	2310076L09Rik
NM_172884	chr5	-	41.98977583	0	1	2900026A02Rik
NM_028725	chr8	-	1.025398667	0	1	4632417N05Rik
NM_028732	chr10	+	10.61635	0	1	4632428N05Rik
NM_172540	chrX	-	29.04042983	0	1	4732479N06Rik
NM_029425	chr3	+	3.226433333	0	1	4833424O15Rik
NM_001078646	chr3	+	2.723946667	0	1	4921515J06Rik
NM_001033170	chr7	+	1.642651	0	1	4930403C10Rik
NM_029053	chr16	-	5.779	0	1	4930451C15Rik
NM_207267	chrX	-	119.235175	0	1	4930488E11Rik
NM_029070	chr16	-	2.51222	0	1	4930511J11Rik
NM_026262	chr11	+	9.746416667	0	1	4930524B15Rik
NM_027644	chr17	-	39.6732	0	1	4931440B09Rik
NM_177230	chr5	-	13.43488333	0	1	4932413O14Rik
NM_175017	chr6	+	13.315773	0	1	4933427D06Rik
NM_021493	chr11	+	38.18616417	0	1	4933428G20Rik
NM_001033304	chr3	-	6.272151667	0	1	5330417C22Rik
NM_001032727	chr15	-	3.617248333	0	1	5730410E15Rik
NM_001033465	chr7	-	11.3524605	0	1	6430531B16Rik
NM_175677	chr7	-	8.526778333	0	1	9230105E10Rik
NM_001081351	chr6	+	5.894446444	0	1	A430107O13Rik
NM_001004174	chr2	+	75.20116667	0	1	AA467197
NM_172961	chr16	+	15.82053333	0	1	Abat
NM_029600	chr11	-	9.345783333	0	1	Abcc3
NM_018795	chr7	-	3.0721	0	1	Abcc6
NM_009599	chr5	+	1.612181333	0	1	Ache
NM_028790	chr13	+	6.214099444	0	1	Acot12
NM_009606	chr8	-	190.4283333	0	1	Acta1
NM_033268	chr13	-	17.38644333	0	1	Actn2

refID	chr	strd	ES_exp	und	3RA	name
NM_172126	chr5	-	9.396541333	0	1	Adam1a
NM_029967	chr4	+	7.953871111	0	1	Adamtsl1
NM_009623	chr15	-	3.796002	0	1	Adcy8
NM_009627	chr7	+	45.18346667	0	1	Adm
NM_013461	chr14	+	3.314106667	0	1	Adra1a
NM_134044	chr12	-	224.184	0	1	AI413782
NM_001039220	chr8	+	27.09473333	0	1	AI429214
NM_027406	chr6	+	24.48371	0	1	Aldh1l1
NM_009662	chr6	-	26.0207	0	1	Alox5
NM_009661	chr11	-	7.138656667	0	1	Alox8
NM_172553	chr10	-	4.674681667	0	1	Alx1
NM_007442	chr2	+	13.95086667	0	1	Alx4
NM_175667	chr2	+	1.466679667	0	1	Ankrd5
NM_019456	chr2	+	11.27054833	0	1	Apbb1ip
NM_013476	chrX	+	4.930932083	0	1	Ar
NM_138630	chrX	-	16.65101111	0	1	Arhgap4
NM_009707	chrX	+	5.040312778	0	1	Arhgap6
NM_007486	chr6	-	38.29646667	0	1	Arhgdib
NM_001081083	chr2	+	3.093956667	0	1	Armc3
NM_001038499	chr18	+	26.35153333	0	1	Arsi
NM_009710	chr7	+	28.72191	0	1	Art1
NM_007495	chr1	+	0.928610667	0	1	Astn1
NM_133699	chr12	-	2.938616667	0	1	Atp6v1c2
NM_015803	chr14	-	10.70687556	0	1	Atp8a2
NM_177809	chr13	+	3.729253667	0	1	AU042651
NM_009732	chr2	-	12.68647333	0	1	Avp
NM_029007	chr12	-	4.273366667	0	1	AW125753
NM_033149	chr16	+	7.256472333	0	1	B3galt5
NM_001005477	chr5	-	8.420356667	0	1	Barhl2
NM_007526	chr13	+	9.090933333	0	1	Barx1
NM_001033128	chr19	-	24.2736	0	1	Bbs1
NM_183162	chr2	-	24.92806667	0	1	BC006779
NM_145389	chr16	-	1.991186667	0	1	BC016579
NM_145577	chr7	-	17.93193	0	1	BC023179
NM_001033284	chr1	+	41.07904667	0	1	BC026585
NM_153584	chr9	+	20.53253889	0	1	BC031353
NM_172295	chr3	-	62.51023333	0	1	BC037703
NM_153172	chr9	+	6.565373333	0	1	BC107230
NM_009738	chr3	-	7.504622444	0	1	Bche
NM_007528	chr11	-	25.4331	0	1	Bcl6b

refID	chr	strd	ES_exp	und	3RA	name
NM_007542	chrX	+	136.4230222	0	1	Bgn
NM_080641	chr2	-	5.18289	0	1	Bhlhb4
NM_022884	chr13	-	5.784758667	0	1	Bhmt2
NM_001079873	chr5	+	12.30290444	0	1	Brdt
NM_133195	chr18	-	28.16772925	0	1	Brunol4
NM_026161	chr2	+	14.95829667	0	1	C1qtnf4
NM_177883	chr6	+	15.46726667	0	1	C330043M08Rik
NM_009782	chr1	-	1.564223333	0	1	Cacna1e
NM_009788	chr4	+	2.224458	0	1	Calb1
NM_021371	chr5	+	3.916604333	0	1	Caln1
NM_028296	chr11	+	7.815518333	0	1	Car10
NM_178396	chr9	+	12.79397478	0	1	Car12
NM_001081493	chr13	-	5.093545333	0	1	Cartpt
NM_013803	chr16	-	9.848468333	0	1	Casr
NM_016900	chr6	+	13.93943333	0	1	Cav2
NM_172633	chr18	+	1.247748	0	1	Cbln2
NM_007626	chr15	-	325.2177753	0	1	Cbx5
NM_172274	chr5	+	53.03453333	0	1	Cc2d2a
NM_146178	chr7	+	41.70396667	0	1	Ccdc106
NM_172914	chr8	+	10.90281667	0	1	Ccdc113
NM_025455	chr4	-	26.6168715	0	1	Ccdc28b
NM_201362	chr18	+	62.61506667	0	1	Ccdc68
NM_001048179	chr4	-	70.69874444	0	1	Ccl27
NM_001039150	chr2	-	38.90918667	0	1	Cd44
NM_009856	chr13	+	3.592613333	0	1	Cd83
NM_028176	chr4	-	32.8504	0	1	Cda
NM_009866	chr8	-	15.796495	0	1	Cdh11
NM_011800	chr1	+	4.06222	0	1	Cdh20
NM_009868	chr8	+	7.93545	0	1	Cdh5
NM_009880	chr18	-	6.198353333	0	1	Cdx1
NM_007674	chrX	+	3.356251	0	1	Cdx4
NM_199032	chr5	+	11.57768	0	1	Cep135
NM_021050	chr6	+	0.553694333	0	1	Cftr
NM_007697	chr6	+	3.436331083	0	1	Chl1
NM_031258	chrX	-	3.15341925	0	1	Chrdl1
NM_033269	chr13	-	6.779846667	0	1	Chrm3
NM_007702	chr18	+	55.0458	0	1	Cidea
NM_173876	chr8	-	327.597275	0	1	Clcn3
NM_172621	chr17	+	13.90686667	0	1	Clic5
NM_153508	chr6	-	45.96013333	0	1	Clstn3

refID	chr	strd	ES_exp	und	3RA	name
NM_001081047	chr4	-	24.799	0	1	Cnksr1
NM_173004	chr6	+	5.875346167	0	1	Cntn4
NM_007729	chr3	+	15.79626667	0	1	Col11a1
NM_007730	chr9	-	13.98062244	0	1	Col12a1
NM_009928	chr4	+	7.367626667	0	1	Col15a1
NM_007733	chr1	-	4.793060667	0	1	Col19a1
NM_007737	chr1	-	32.95301283	0	1	Col5a2
NM_016919	chr9	-	19.43762833	0	1	Col5a3
NM_146007	chr10	-	26.69700667	0	1	Col6a2
NM_199473	chr4	+	11.10308333	0	1	Col8a2
NM_009898	chr7	-	42.26711444	0	1	Coro1a
NM_030703	chr19	-	19.41252	0	1	Cpn1
NM_007758	chr1	-	2.665696667	0	1	Cr2
NM_007759	chr3	+	191.5546667	0	1	Crabp2
NM_030209	chr8	+	11.88480889	0	1	Crispld2
NM_021541	chr1	-	12.27302	0	1	Cryba2
NM_145473	chr15	+	15.80996444	0	1	Csdc2
NM_139001	chr9	+	4.572242	0	1	Cspg4
NM_026778	chr15	+	51.1359	0	1	Cthrc1
NM_001012477	chr6	+	30.96568333	0	1	Cxcl12
NM_009140	chr5	+	5.13467	0	1	Cxcl2
NM_009996	chr2	-	6.758166667	0	1	Cyp24a1
NM_010008	chr4	-	5.051503333	0	1	Cyp2j6
NM_177307	chr17	+	36.02816667	0	1	Cyp4f39
NM_001003947	chr4	-	6.350996667	0	1	Cyp4x1
NM_001024931	chr11	-	15.48262	0	1	D11Bwg0517e
NM_001081051	chr13	+	5.68623	0	1	D130043K22Rik
NM_153574	chr6	-	8.775338333	0	1	D430015B01Rik
NM_033079	chr6	+	51.733	0	1	D6Mm5e
NM_001008231	chr17	-	6.258941833	0	1	Daam2
NM_001037905	chr15	+	141.4350167	0	1	Dab2
NM_001001602	chr2	+	103.1778667	0	1	Dab2ip
NM_001005232	chr7	-	5.168793333	0	1	Dbx1
NM_207533	chr15	-	0.959807333	0	1	Dbx2
NM_007835	chr6	+	220.4691667	0	1	Dctn1
NM_177914	chrX	+	7.97494	0	1	Dgkk
NM_020265	chr3	+	2.207786333	0	1	Dkk2
NM_015789	chr7	-	26.21506667	0	1	Dkkl1
NM_010052	chr12	+	25.89073333	0	1	Dlk1
NM_010055	chr11	+	14.53574	0	1	Dlx3

refID	chr	strd	ES_exp	und	3RA	name
NM_028772	chr13	+	6.908885	0	1	Dmgdh
NM_145831	chr19	+	2.971846667	0	1	Dmrt2
NM_001034878	chr11	+	15.374675	0	1	Dnaic2
NM_025926	chr3	-	33.89873267	0	1	Dnajb4
NM_001038619	chr1	-	9.031808556	0	1	Dnm3
NM_021791	chr19	+	43.40593333	0	1	Doc2g
NM_010071	chr14	+	142.5763333	0	1	Dok2
NM_199021	chr1	-	3.335886722	0	1	Dpp10
NM_007878	chr7	+	9.339783333	0	1	Drd4
NM_007882	chr18	-	5.281264333	0	1	Dsc3
NM_025777	chr2	+	10.55824667	0	1	Duoxa2
NM_001077694	chr6	+	46.065285	0	1	Dysf
NM_010095	chr14	+	4.100060333	0	1	Ebf2
NM_021306	chr1	-	18.14236667	0	1	Ecel1
NM_007903	chr2	+	11.35396583	0	1	Edn3
NM_010332	chr8	-	4.8863125	0	1	Ednra
NM_146015	chr11	+	25.9915	0	1	Efemp1
NM_001034882	chr14	+	12.73593333	0	1	EG432879
NM_153068	chr7	-	18.00237833	0	1	Ehd2
NM_023737	chr16	-	8.27246	0	1	Ehhadh
NM_183141	chr15	-	9.98515	0	1	Elfn2
NM_024474	chr5	-	3.308851667	0	1	Emid2
NM_133918	chr5	+	7.10736	0	1	Emilin1
NM_010131	chr6	+	4.887960833	0	1	Emx1
NM_010132	chr19	+	5.7946	0	1	Emx2
NM_010133	chr1	+	4.57889	0	1	En1
NM_007933	chr11	+	78.660805	0	1	Eno3
NM_032003	chr17	+	75.98556667	0	1	Enpp5
NM_009848	chr19	+	3.140346667	0	1	Entpd1
NM_023580	chr6	-	15.94314889	0	1	Epha1
NM_027984	chr11	-	2.8445	0	1	Epn3
NM_029495	chr14	+	8.528187667	0	1	Epsti1
NM_010154	chr1	-	4.213521333	0	1	Erbb4
NM_007950	chr5	+	14.50606667	0	1	Ereg
NM_023154	chr7	+	134.0493333	0	1	Ethe1
NM_007959	chr7	-	46.64126667	0	1	Etv2
NM_007966	chr6	+	10.18014333	0	1	Evx1
NM_015795	chr14	+	37.87736667	0	1	Fbxo16
NM_153111	chr1	-	10.46433333	0	1	Fev
NM_030610	chr8	-	3.965891333	0	1	Fgf20

refID	chr	strd	ES_exp	und	3RA	name
NM_008034	chr7	-	405.341	0	1	Folr1
NM_008259	chr12	-	6.605289667	0	1	Foxa1
NM_008023	chr19	-	1.889806667	0	1	Foxb2
NM_008592	chr13	+	10.06806	0	1	Foxc1
NM_010225	chr13	+	7.128838	0	1	Foxf2
NM_010237	chr10	+	6.323597133	0	1	Frk
NM_172869	chr4	+	5.065050833	0	1	Frmd3
NM_172475	chr2	+	62.563515	0	1	Frmd4a
NM_177136	chr5	-	99.041375	0	1	Fryl
NM_178673	chr3	+	6.596066667	0	1	Fstl5
NM_001081454	chr7	-	88.59076667	0	1	Furin
NM_019503	chr7	-	19.90463333	0	1	Fxyd1
NM_022007	chr7	-	2.891096	0	1	Fxyd7
NM_008067	chrX	-	4.538809333	0	1	Gabra3
NM_008070	chr11	+	9.478448667	0	1	Gabrb2
NM_010252	chr5	-	3.089753667	0	1	Gabrg1
NM_173739	chr7	-	13.74223333	0	1	Galntl4
NM_008083	chr16	-	10.750891	0	1	Gap43
NM_010258	chr18	+	45.39638333	0	1	Gata6
NM_001077411	chr3	+	174.8891667	0	1	Gba
NM_018734	chr3	+	43.25273333	0	1	Gbp3
NM_010267	chr1	+	8.4745415	0	1	Gdap1
NM_145741	chr14	+	5.137793333	0	1	Gdf10
NM_010278	chr5	-	2.519437	0	1	Gfi1
NM_010280	chr18	-	34.28203333	0	1	Gfra3
NM_144786	chr2	-	2.382863333	0	1	Ggt7
NM_001010937	chr14	-	5.285336667	0	1	Gjb6
NM_010290	chr2	-	4.26456	0	1	Gjd2
NM_025374	chr17	-	753.5391	0	1	Glo1
NM_001085513	chr4	+	29.57993333	0	1	Gm693
NM_001007580	chrX	-	81.5337	0	1	Gm784
NM_001005424	chr2	-	6.164707	0	1	Gm996
NM_025331	chr6	+	6.849093333	0	1	Gng11
NM_026680	chr1	+	6.171436667	0	1	Golt1a
NM_198192	chr3	-	4.149433333	0	1	Gpr103
NM_177346	chr3	-	1.293719222	0	1	Gpr149
NM_173365	chr15	-	33.82423333	0	1	Gpr20
NM_001079847	chrX	+	2.820535	0	1	Gpr64
NM_030720	chr15	-	9.4807	0	1	Gpr84
NM_182805	chr15	+	55.45773333	0	1	Gpt

refID	chr	strd	ES_exp	und	3RA	name
NM_153528	chr16	-	4.410622667	0	1	Gramd1c
NM_001039195	chr3	-	8.472208833	0	1	Gria2
NM_016886	chrX	+	2.503239833	0	1	Gria3
NM_133355	chr5	+	20.13223333	0	1	Grid2ip
NM_010348	chr16	-	1.687404833	0	1	Grik1
NM_010350	chr11	-	5.344683333	0	1	Grin2c
NM_181850	chr5	-	3.13942	0	1	Grm3
NM_177328	chr6	+	3.185946833	0	1	Grm7
NM_001080553	chr6	-	12.52174	0	1	Gsg1
NM_133256	chr5	+	8.280863333	0	1	Gsx2
NM_023630	chr17	+	6.832416667	0	1	Gtf2a1l
NM_021896	chr3	-	5.387448	0	1	Gucy1a3
NM_017469	chr3	-	18.75116667	0	1	Gucy1b3
NM_178747	chr14	-	24.01483333	0	1	Gulo
NM_010369	chr8	+	2.144092833	0	1	Gypa
NM_001048207	chr18	-	16.9261	0	1	Gypc
NM_013500	chr13	+	4.121338722	0	1	Hapln1
NM_009330	chr11	+	12.29476667	0	1	Hnf1b
NM_013920	chr3	+	2.258252111	0	1	Hnf4g
NM_008263	chr6	-	4.802387833	0	1	Hoxa10
NM_010450	chr6	-	14.3211	0	1	Hoxa11
NM_010451	chr6	-	6.729223333	0	1	Hoxa2
NM_010455	chr6	-	3.9294	0	1	Hoxa7
NM_010456	chr6	-	16.056205	0	1	Hoxa9
NM_008267	chr11	+	5.42563	0	1	Hoxb13
NM_134032	chr11	+	16.10728667	0	1	Hoxb2
NM_010458	chr11	+	10.03229667	0	1	Hoxb3
NM_010459	chr11	+	7.950618333	0	1	Hoxb4
NM_008268	chr11	+	1.575192333	0	1	Hoxb5
NM_008269	chr11	+	3.263401	0	1	Hoxb6
NM_010460	chr11	+	5.128853333	0	1	Hoxb7
NM_010461	chr11	+	7.116043333	0	1	Hoxb8
NM_013553	chr15	+	2.986306667	0	1	Hoxc4
NM_175730	chr15	+	2.922351556	0	1	Hoxc5
NM_010466	chr15	+	4.701613333	0	1	Hoxc8
NM_013554	chr2	+	3.160286667	0	1	Hoxd10
NM_010468	chr2	+	5.093916667	0	1	Hoxd3
NM_013555	chr2	+	8.84794	0	1	Hoxd9
NM_152803	chr5	-	16.16563333	0	1	Hpse
NM_001045527	chr11	+	22.51836667	0	1	Hsf5

refID	chr	strd	ES_exp	und	3RA	name
NM_028306	chr2	+	1.444075667	0	1	Hspa12b
NM_010483	chr1	-	5.209976667	0	1	Htr5b
NM_001042489	chr5	+	82.3278	0	1	Hvcn1
NM_023892	chr9	+	19.51003333	0	1	Icam4
NM_027835	chr2	-	6.151863333	0	1	Ifih1
NM_028968	chr16	-	15.35856667	0	1	Ifitm7
NM_010512	chr10	+	11.59969422	0	1	Igf1
NM_010514	chr7	-	83.30496667	0	1	Igf2
NM_010544	chr1	-	10.60878333	0	1	Ihh
NM_011771	chr11	-	6.374703	0	1	Ikzf3
NM_008357	chr8	-	8.35968	0	1	Il15
NM_008358	chr2	+	5.863516667	0	1	Il15ra
NM_013563	chrX	-	26.68428333	0	1	Il2rg
NM_001008700	chr7	+	5.956526778	0	1	Il4ra
NM_008380	chr13	+	30.73955833	0	1	Inhba
NM_013754	chr19	-	53.3247	0	1	Insl6
NM_011832	chr3	+	6.265636667	0	1	Insrr
NM_016850	chr7	-	4.453696667	0	1	Irf7
NM_010572	chrX	-	15.542355	0	1	Irs4
NM_010573	chr13	-	12.77979667	0	1	Irx1
NM_018885	chr13	+	1.703066667	0	1	Irx4
NM_172471	chr2	+	2.003182	0	1	Itih5
NM_001017426	chr11	-	189.7974833	0	1	Jmjd3
NM_021566	chr2	-	10.79746667	0	1	Jph2
NM_010590	chr14	-	963.1553333	0	1	Jub
NM_027721	chr18	-	17.28128333	0	1	Katnal2
NM_001024135	chr14	+	64.93765911	0	1	Kbtbd7
NM_013568	chr6	-	5.748365556	0	1	Kcna6
NM_021487	chrX	-	2.997803333	0	1	Kcne1l
NM_201531	chr12	-	2.652526833	0	1	Kcnf1
NM_027398	chr11	-	9.952791667	0	1	Kcnip1
NM_001039484	chr1	+	15.69733333	0	1	Kcnj10
NM_010638	chr19	+	767.4868529	0	1	Klf9
NM_053105	chr14	-	2.916557	0	1	Klhl1
NM_177261	chr7	+	20.47016556	0	1	Kndc1
NM_145416	chr9	-	81.58837917	0	1	Kri1
NM_010664	chr15	+	603.3416667	0	1	Krt18
NM_023256	chr11	-	30.07633333	0	1	Krt20
NM_172946	chr11	-	3.892977	0	1	Krt222
NM_033073	chr15	+	87.78863333	0	1	Krt7

refID	chr	strd	ES_exp	und	3RA	name
NM_010681	chr10	+	2.539671	0	1	Lama4
NM_010692	chr6	+	13.874	0	1	Lbx2
NM_010701	chr14	-	2.172976667	0	1	Lect1
NM_008495	chr15	+	1444.747333	0	1	Lgals1
NM_010708	chr11	-	65.76143333	0	1	Lgals9
NM_145219	chr14	+	5.650827444	0	1	Lgi3
NM_023063	chr15	-	417.3001421	0	1	Lima1
NM_053106	chr1	+	8.558926667	0	1	Lmod1
NM_001024712	chr14	-	610.4124111	0	1	LOC544988
NM_001081983	chrX	-	3.50135	0	1	LOC666244
NM_134152	chr19	+	3.216616667	0	1	Lpxn
NM_019457	chr15	-	3.063235167	0	1	Lrrc6
NM_025730	chr15	+	7.4304555	0	1	Lrrk2
NM_001013019	chr19	+	27.37016667	0	1	Lrrn4cl
NM_028880	chr6	+	2.143821	0	1	Lrrtm1
NM_172492	chr6	-	1.924463333	0	1	Lrtm2
NM_175548	chr16	+	4.576220889	0	1	Lsamp
NM_010736	chr6	-	37.55273333	0	1	Ltbr
NM_178705	chr7	+	0.645647333	0	1	Luzp2
NM_023463	chr17	+	16.99554	0	1	Ly6g6c
NM_145532	chr2	-	7.322345833	0	1	Mall
NM_174857	chr19	-	2.401620667	0	1	Mamdc2
NM_001081354	chrX	+	16.51658	0	1	Mamld1
NM_008555	chr16	-	10.9115836	0	1	Masp1
NM_013592	chr2	-	7.85363	0	1	Matn4
NM_001025251	chr18	+	19.4648255	0	1	Mbp
NM_181407	chr7	+	6.315526667	0	1	Me3
NM_001001979	chr18	+	2.130553333	0	1	Megf10
NM_008595	chr15	-	63.85243333	0	1	Mfng
NM_029662	chr4	-	9.178802333	0	1	Mfsd2
NM_172499	chr1	-	19.36157167	0	1	Mfsd9
NM_153049	chr15	-	79.07283333	0	1	Mkl1
NM_198831	chr7	-	417.3805	0	1	Mrpl48
NM_019544	chr12	-	13.78201	0	1	Msgn1
NM_181409	chr3	+	17.53432778	0	1	Mtmr11
NM_008639	chr8	+	6.679793333	0	1	Mtnr1a
NM_001005865	chr8	-	33.09148667	0	1	Mtus1
NM_010850	chrX	-	6.77478	0	1	Mycs
NM_145136	chr11	-	4.554471167	0	1	Myocd
NM_011848	chr8	-	137.122	0	1	Nek3

refID	chr	strd	ES_exp	und	3RA	name
NM_001037906	chr7	+	3.845489667	0	1	Nell1
NM_144946	chr18	+	3.6237675	0	1	Neto1
NM_010896	chr13	-	9.241877333	0	1	Neurog1
NM_009719	chr10	+	8.037597167	0	1	Neurog3
NM_001037178	chr2	-	13.50962111	0	1	Nfatc2
NM_008685	chr15	-	2.827183333	0	1	Nfe2
NM_022414	chr12	-	8.56494	0	1	Ngb
NM_178777	chr3	+	2.919561667	0	1	Nhlh2
NM_009385	chr12	-	1.642526667	0	1	Nkx2-1
NM_001077632	chr2	-	3.752078667	0	1	Nkx2-2
NM_008699	chr19	+	11.15045111	0	1	Nkx2-3
NM_008700	chr17	-	25.90706667	0	1	Nkx2-5
NM_007524	chr5	-	2.467476667	0	1	Nkx3-2
NM_138666	chr3	-	3.185986333	0	1	Nlgn1
NM_199024	chr18	-	2.965663778	0	1	Nol4
NM_027988	chr17	+	4.72214	0	1	Noxo1
NM_013780	chr12	+	6.081867417	0	1	Npas3
NM_130456	chr1	+	3.697845667	0	1	Nphs2
NM_008726	chr4	+	38.28886667	0	1	Nppb
NM_010934	chr8	+	7.469448333	0	1	Npy1r
NM_016708	chr8	-	2.9322395	0	1	Npy5r
NM_010151	chr13	-	5.292623333	0	1	Nr2f1
NM_008173	chr18	-	45.51891853	0	1	Nr3c1
NM_015743	chr4	+	3.752434444	0	1	Nr4a3
NM_008734	chr14	-	5.228855952	0	1	Nrg3
NM_013724	chrX	+	8.337085556	0	1	Nrk
NM_009513	chr13	-	29.0076	0	1	Nrsn1
NM_018766	chr2	+	4.94512	0	1	Ntsr1
NM_021431	chrX	+	42.75136667	0	1	Nudt11
NM_029385	chr9	-	38.93716667	0	1	Nudt16
NM_031259	chrX	-	14.4451	0	1	Nxf2
NM_145211	chr5	-	78.6007	0	1	Oas1a
NM_177068	chr1	+	17.54806667	0	1	Olfml2b
NM_133859	chr3	-	9.527223333	0	1	Olfml3
NM_147049	chr7	-	2.254426667	0	1	Olfr658
NM_016967	chr16	+	177.9406667	0	1	Olig2
NM_139226	chr10	+	1.587215	0	1	Onecut3
NM_011012	chr2	+	6.620716556	0	1	Oprl1
NM_054076	chr1	-	7.088313333	0	1	Optc
NM_024289	chr7	-	14.06153333	0	1	Osbpl5

refID	chr	strd	ES_exp	und	3RA	name
NM_001081147	chr6	-	4.429875	0	1	Oxtr
NM_011028	chr16	+	10.71542	0	1	P2rx6
NM_130887	chr12	+	0.855134	0	1	Papln
NM_027995	chr4	+	14.803495	0	1	Paqr7
NM_008780	chr2	+	2.171236667	0	1	Pax1
NM_025273	chr10	+	153.947	0	1	Pcbd1
NM_001081385	chrX	+	1.183087	0	1	Pcdh11x
NM_001013753	chr14	+	1.699420889	0	1	Pcdh17
NM_054072	chr18	+	4.014749889	0	1	Pcdha1
NM_138662	chr18	+	4.014749889	0	1	Pcdha3
NM_009959	chr18	+	4.014749889	0	1	Pcdha5
NM_009957	chr18	+	4.014749889	0	1	Pcdha7
NM_138661	chr18	+	4.014749889	0	1	Pcdha9
NM_053126	chr18	+	3.731304333	0	1	Pcdhb1
NM_053139	chr18	+	1.863453	0	1	Pcdhb14
NM_053146	chr18	+	2.617686333	0	1	Pcdhb21
NM_053131	chr18	+	1.102107667	0	1	Pcdhb6
NM_033579	chr18	+	170.787	0	1	Pcdhgb7
NM_153565	chr4	-	24.70686667	0	1	Pcsk9
NM_001081033	chr2	-	4.127126667	0	1	Pde11a
NM_001039376	chr3	-	37.4920075	0	1	Pde4dip
NM_027924	chr9	+	6.301365333	0	1	Pdgfd
NM_011058	chr5	+	41.70233444	0	1	Pdgfra
NM_016798	chr8	+	23.81056667	0	1	Pdlim3
NM_019417	chr11	-	66.39936667	0	1	Pdlim4
NM_001013369	chr2	+	1511.94	0	1	Pfdn4
NM_029303	chr13	-	15.24156667	0	1	Pfn3
NM_008827	chr12	-	16.23056667	0	1	Pgf
NM_175013	chr19	-	1.477770333	0	1	Pgm5
NM_008840	chr4	-	50.59018133	0	1	Pik3cd
NM_011097	chr13	-	6.540935	0	1	Pitx1
NM_001042504	chr3	+	268.802	0	1	Pitx2
NM_001039051	chr10	+	1.249808167	0	1	Pkib
NM_019645	chr1	-	7.2576	0	1	Pkp1
NM_008869	chr1	-	12.84538333	0	1	Pla2g4a
NM_207279	chr5	+	15.96748	0	1	Plcxd1
NM_013738	chr12	-	28.25993333	0	1	Plek2
NM_027196	chr19	+	130.3963333	0	1	Pold4
NM_008901	chrX	+	5.299595833	0	1	Pou3f4
NM_001080963	chr7	+	14.08566667	0	1	Ppapdc1a

refID	chr	strd	ES_exp	und	3RA	name
NM_011169	chr15	+	6.281253333	0	1	Prlr
NM_178774	chr17	+	3.777096667	0	1	Prr18
NM_028707	chr18	+	1.457858667	0	1	Psd2
NM_001093750	chrX	-	10.46040333	0	1	Ptchd1
NM_013641	chr8	+	23.69642833	0	1	Ptger1
NM_053256	chr7	+	26.4975	0	1	Pth2
NM_008970	chr6	-	3.1040055	0	1	Pthlh
NM_025760	chr4	-	23.75793	0	1	Ptplad2
NM_027900	chr10	+	213.7746667	0	1	R3hdm2
NM_026677	chr3	+	88.00843333	0	1	Rab13
NM_198409	chrX	+	1.645422667	0	1	Rai2
NM_001033158	chr9	+	19.4411	0	1	Rasl12
NM_009028	chr7	-	9.147291667	0	1	Rasl2-9
NM_028478	chr5	-	39.41113333	0	1	Rassf6
NM_013833	chr18	-	9.926003333	0	1	Rax
NM_011254	chr9	+	107.6246333	0	1	Rbp1
NM_001080928	chr5	+	794.4473492	0	1	Rbpj
NM_009036	chr2	+	2.620941667	0	1	Rbpjl
NM_030017	chr12	+	5.16908	0	1	Rdh12
NM_020002	chr14	+	100.9974	0	1	Rec8
NM_001024918	chr10	+	4.11309	0	1	Rfx4
NM_023622	chr9	-	20.77326333	0	1	Rgl3
NM_019799	chr7	-	2.532366667	0	1	Rhcg
NM_023275	chr12	+	10.82528833	0	1	Rhoj
NM_023894	chrX	-	79.9963	0	1	Rhox9
NM_001081388	chr5	-	17.40011333	0	1	Rimbp2
NM_028724	chr2	+	11.31981	0	1	Rin2
NM_001083938	chr17	-	1287.257667	0	1	Rnaset2a
NM_026611	chr17	-	862.8054678	0	1	Rnaset2b
NM_026205	chr17	-	17.66623333	0	1	Rnf151
NM_001033489	chr4	-	30.76783333	0	1	Rnf207
NM_001043354	chr19	-	6.738308333	0	1	Rorb
NM_026594	chr16	+	573.852	0	1	Rpl39l
NM_011302	chrX	+	13.43143167	0	1	Rs1
NM_028351	chr10	-	2.891793333	0	1	Rspo3
NM_001025364	chr7	+	60.73233333	0	1	Rtn2
NM_009107	chr1	+	15.12421667	0	1	Rxrg
NM_010101	chr13	+	6.578155556	0	1	S1pr3
NM_172809	chr14	+	72.55737822	0	1	Sacs
NM_175303	chr2	-	1050.996833	0	1	Sall4

refID	chr	strd	ES_exp	und	3RA	name
NM_146025	chr11	+	22.5558	0	1	Samd14
NM_153790	chr16	+	11.86023922	0	1	Scarf2
NM_018852	chr2	-	3.954262444	0	1	Scn9a
NM_009150	chr3	+	33.753	0	1	Selenbp1
NM_011340	chr11	-	78.93481667	0	1	Serpinf1
NM_009776	chr2	-	146.3313333	0	1	Serping1
NM_011891	chr11	-	3.509288833	0	1	Sgcd
NM_015825	chr16	+	6.589544889	0	1	Sh3bgr
NM_001077595	chr5	+	80.991668	0	1	Shroom3
NM_001033186	chr11	+	3.391901714	0	1	Skap1
NM_173403	chr5	+	2.790973667	0	1	Slc10a4
NM_009197	chrX	-	16.65259167	0	1	Slc16a2
NM_020516	chr15	-	19.35266667	0	1	Slc16a8
NM_009200	chr10	+	4.526708667	0	1	Slc1a6
NM_011395	chr17	-	4.803183333	0	1	Slc22a3
NM_019687	chr11	-	3.779623333	0	1	Slc22a4
NM_178934	chr10	+	9.328763333	0	1	Slc2a12
NM_022025	chr17	-	8.678536667	0	1	Slc5a7
NM_145423	chr10	+	5.597236667	0	1	Slc5a8
NM_201353	chr18	-	25.54696667	0	1	Slc6a7
NM_001033289	chr1	+	3.62139	0	1	Slc9a2
NM_198864	chr3	-	2.44498	0	1	Slitrk3
NM_178740	chrX	-	7.028866667	0	1	Slitrk4
NM_026003	chr19	+	8.507941111	0	1	Smarca2
NM_011427	chr2	+	17.72363333	0	1	Snai1
NM_025696	chr19	+	8.51482	0	1	Sorcs3
NM_001025428	chr11	+	146.9288078	0	1	Spag9
NM_027641	chr2	-	9.264008333	0	1	Spef1
NM_007463	chr1	+	22.55948333	0	1	Speg
NM_172430	chr1	-	2.980400778	0	1	Sphkap
NM_145584	chr7	+	8.204268889	0	1	Spon1
NM_080448	chr6	-	32.6975	0	1	Srgap3
NM_011371	chr11	-	1.333303333	0	1	St6galnac1
NM_012028	chr3	-	5.741171333	0	1	St6galnac5
NM_001083315	chr6	+	9.88613	0	1	St7
NM_029682	chr19	+	39.92205	0	1	Stambpl1
NM_145934	chr17	-	149.3431667	0	1	Stap2
NM_022416	chr5	-	10.22711833	0	1	Stk32b
NM_009292	chr6	+	147.239	0	1	Stra8
NM_025491	chr13	-	4.905306667	0	1	Susd3

refID	chr	strd	ES_exp	und	3RA	name
NM_022030	chr3	+	29.6678	0	1	Sv2a
NM_001079686	chr10	+	13.89172222	0	1	Syne1
NM_018804	chr3	-	23.81815167	0	1	Syt11
NM_030725	chr2	+	20.34911667	0	1	Syt13
NM_016663	chr7	+	20.63093333	0	1	Syt3
NM_018800	chr3	+	4.507696	0	1	Syt6
NM_009309	chr17	+	26.5558	0	1	T
NM_009311	chr6	+	7.638955667	0	1	Tac1
NM_009313	chr6	+	1.508643333	0	1	Tacr1
NM_020047	chr6	-	8.873333333	0	1	Tacstd2
NM_028958	chrX	-	18.0802	0	1	Taf7l
NM_145391	chr6	-	12.92504833	0	1	Tapbpl
NM_009322	chr2	+	6.609723333	0	1	Tbr1
NM_011532	chr16	-	1.018372333	0	1	Tbx1
NM_146236	chrX	+	34.31716667	0	1	Tceal1
NM_001029978	chrX	+	2.31834	0	1	Tceal3
NM_011547	chr13	-	16.13752111	0	1	Tcfap2a
NM_001025305	chr1	+	11.33709189	0	1	Tcfap2b
NM_015749	chr11	-	167.29026	0	1	Tcn2
NM_153533	chr15	+	10.32118667	0	1	Tenc1
NM_031381	chrX	-	6.4734265	0	1	Tex13
NM_009364	chr6	-	10.792	0	1	Tfpi2
NM_009369	chr13	+	99.73055	0	1	Tgfbi
NM_011580	chr2	+	558.5631667	0	1	Thbs1
NM_009381	chr7	-	19.89140333	0	1	Thrsp
NM_028927	chr8	+	3.98509	0	1	Tktl2
NM_011904	chr19	-	0.618975	0	1	Tll2
NM_126166	chr8	-	4.988200667	0	1	Tlr3
NM_001040130	chr2	-	67.82738333	0	1	Tmem141
NM_178642	chr7	-	3.657576333	0	1	Tmem16a
NM_177794	chr10	+	3.843107333	0	1	Tmem26
NM_009396	chr12	+	22.19111778	0	1	Tnfaip2
NM_025566	chr17	+	59.07876667	0	1	Tnfaip8l1
NM_028075	chr15	-	31.37126667	0	1	Tnfrsf13c
NM_011610	chr4	-	50.70891667	0	1	Tnfrsf1b
NM_024290	chr7	-	27.38840333	0	1	Tnfrsf23
NM_021327	chr11	-	718.3313333	0	1	Tnip1
NM_001081260	chr2	+	59.51983333	0	1	Tnks1bp1
NM_031176	chr17	+	6.205793333	0	1	Tnxb
NM_145711	chr4	-	4.990573667	0	1	Tox

refID	chr	strd	ES_exp	und	3RA	name
NM_001024134	chr17	-	9.0054	0	1	Trim15
NM_177781	chr1	-	8.464893333	0	1	Trpa1
NM_080455	chr2	+	4.311494583	0	1	Tshz2
NM_001083618	chr2	+	5.252902111	0	1	Ttll9
NM_009451	chr17	-	23.6571	0	1	Tubb4
NM_011656	chr3	-	42.18440667	0	1	Tuft1
NM_175309	chr5	+	3.055062833	0	1	Upk3b
NM_011909	chr6	+	7.571813333	0	1	Usp18
NM_181399	chr2	+	98.48841667	0	1	Usp6nl
NM_011691	chr17	+	3.3373	0	1	Vav1
NM_009501	chr19	-	1.219994	0	1	Vax1
NM_177683	chr6	-	24.1832	0	1	Vgll4
NM_011700	chr9	+	107.5786333	0	1	Vill
NM_012038	chr12	-	3.597153	0	1	Vsnl1
NM_027725	chr1	+	1.119412667	0	1	Wdr69
NM_021279	chr15	+	3.76374	0	1	Wnt1
NM_053116	chr6	+	3.84006	0	1	Wnt16
NM_009521	chr11	+	3.233383333	0	1	Wnt3
NM_133208	chr11	-	12.7194575	0	1	Zfp287
NM_172385	chr7	-	5.634516667	0	1	Zfp536
NM_001024950	chr17	+	20.98453333	0	1	Zfp563
NM_001033205	chr7	-	13.39923667	0	1	Zfp575
NM_177622	chr13	-	31.55963333	0	1	Zfp595
NM_175751	chr18	-	19.52446333	0	1	Zfp608
NM_172486	chr17	+	43.14073333	0	1	Zfp677
NM_153194	chr15	+	310.272	0	1	Zfp740
NM_175513	chr2	+	1.281704167	0	1	Zfp804a
NM_009573	chr9	-	2.518656667	0	1	Zic1
NM_177086	chr8	+	8.355600778	0	1	Zmat4
NM_011483	chr17	-	1.296146667	0	1	Znrf4

References

Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell *9*, 1191-1200.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Molecular Biology of the Cell, 4th edn.

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

Allfrey, V.G., Littau, V.C., and Mirsky, A.E. (1963). On the role of of histones in regulation ribonucleic acid synthesis in the cell nucleus. Proc Natl Acad Sci U S A *49*, 414-421.

Allfrey, V.G., and Mirsky, A.E. (1962). Evidence for the complete DNA-dependence of RNA synthesis in isolated thymus nuclei. Proc Natl Acad Sci U S A *48*, 1590-1596.

Allis, C.D., Bowen, J.K., Abraham, G.N., Glover, C.V., and Gorovsky, M.A. (1980). Proteolytic processing of histone H3 in chromatin: a physiologically regulated event in Tetrahymena micronuclei. Cell *20*, 55-64.

Allis, C.D., Jenuwien, T., and Reinberg, D. (2006). Epigenetics (Cold Spring Harbor Laboratory Press).

Anand, R., and Marmorstein, R. (2007). Structure and mechanism of lysine-specific demethylase enzymes. J Biol Chem *282*, 35425-35429.

Arney, K.L., and Fisher, A.G. (2004). Epigenetic aspects of differentiation. J Cell Sci *117*, 4355-4363.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M.*, et al.* (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol *8*, 532-538.

Badenhorst, P., Xiao, H., Cherbas, L., Kwon, S.Y., Voas, M., Rebay, I., Cherbas, P., and Wu, C. (2005). The Drosophila nucleosome remodeling factor NURF is required for Ecdysteroid signaling and metamorphosis. Genes Dev *19*, 2540-2545.

Bannister, A.J., and Kouzarides, T. (2004). Histone methylation: recognizing the methyl mark. Methods Enzymol *376*, 269-288.

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

Barr, M.L., and Bertram, E.G. (1949). A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. Nature *163*, 676.

Barrett, A.J., Kembhavi, A.A., Brown, M.A., Kirschke, H., Knight, C.G., Tamai, M., and Hanada, K. (1982). L-trans-Epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L. Biochem J *201*, 189-198.

Barrett, A.J., Rawlings, N.D., and Woessner, J.F., eds. (2003). Handbook of Proteolytic Enzymes, 2 addition edn (Academic Press).

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

Bench, A.J., Nacheva, E.P., Hood, T.L., Holden, J.L., French, L., Swanton, S., Champion, K.M., Li, J., Whittaker, P., Stavrides, G.*, et al.* (2000). Chromosome 20 deletions in myeloid malignancies: reduction of the common deleted region, generation of a PAC/BAC contig and identification of candidate genes. UK Cancer Cytogenetics Group (UKCCG). Oncogene *19*, 3902-3913.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006a). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bernstein, E., and Allis, C.D. (2005). RNA meets chromatin. Genes Dev *19*, 1635-1655.

Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., and Allis, C.D. (2006b). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol Cell Biol *26*, 2560-2569.

Bloom, K.S., and Anderson, J.N. (1978). Fractionation and characterization of chromosomal proteins by the hydroxyapatite dissociation method. J Biol Chem *253*, 4446-4450.

Boccuni, P., MacGrogan, D., Scandura, J.M., and Nimer, S.D. (2003). The human L(3)MBT polycomb group protein is a transcriptional repressor and interacts physically and functionally with TEL (ETV6). J Biol Chem *278*, 15412-15420.

Bornemann, D., Miller, E., and Simon, J. (1996). The Drosophila Polycomb group gene Sex comb on midleg (Scm) encodes a zinc finger protein with similarity to polyhomeotic protein. Development *122*, 1621-1630.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K.*, et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev *20*, 1123-1136.

Bracken, A.P., Kleine-Kohlbrecher, D., Dietrich, N., Pasini, D., Gargiulo, G., Beekman, C., Theilgaard-Monch, K., Minucci, S., Porse, B.T., Marine, J.C.*, et al.* (2007). The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. Genes Dev *21*, 525-530.

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

Bulynko, Y.A., Hsing, L.C., Mason, R.W., Tremethick, D.J., and Grigoryev, S.A. (2006). Cathepsin L stabilizes the histone modification landscape on the Y chromosome and pericentromeric heterochromatin. Mol Cell Biol *26*, 4172-4184.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-1043.

Chaumeil, J., Okamoto, I., Guggiari, M., and Heard, E. (2002). Integrated kinetics of X chromosome inactivation in differentiating embryonic stem cells. Cytogenet Genome Res *99*, 75-84.

Chow, C.M., Georgiou, A., Szutorisz, H., Maia e Silva, A., Pombo, A., Barahona, I., Dargelos, E., Canzonetta, C., and Dillon, N. (2005). Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. EMBO Rep *6*, 354-360.

Conliffe, P.R., Ogilvie, S., Simmen, R.C., Michel, F.J., Saunders, P., and Shiverick, K.T. (1995). Cloning and expression of a rat placental cDNA encoding a novel cathepsin L-related protein. Mol Reprod Dev *40*, 146-156.

Coon, J.J., Ueberheide, B., Syka, J.E., Dryhurst, D.D., Ausio, J., Shabanowitz, J., and Hunt, D.F. (2005). Protein identification using sequential ion/ion reactions and tandem mass spectrometry. Proc Natl Acad Sci U S A *102*, 9463-9468.

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

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell *111*, 185-196.

Daury, L., Chailleux, C., Bonvallet, J., and Trouche, D. (2006). Histone H3.3 deposition at E2F-regulated genes is linked to transcription. EMBO Rep *7*, 66-71.

de Napoles, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M.*, et al.* (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev Cell *7*, 663-676.

de Ruijter, A.J., van Gennip, A.H., Caron, H.N., Kemp, S., and van Kuilenburg, A.B. (2003). Histone deacetylases (HDACs): characterization of the classical HDAC family. Biochem J *370*, 737-749.

DeSouza, C.A., Shapiro, L.F., Clevenger, C.M., Dinenno, F.A., Monahan, K.D., Tanaka, H., and Seals, D.R. (2000). Regular aerobic exercise prevents and restores age-related declines in endothelium-dependent vasodilation in healthy men. Circulation *102*, 1351-1357.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491-496.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-1489.

Duncan, E.M., Muratore-Schroeder, T.L., Cook, R.G., Garcia, B.A., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2008). Cathepsin L Proteolytically Processes Histone H3 During Mouse Embryonic Stem Cell Differentiation. Cell *135*, 284-294.

Durrin, L.K., Mann, R.K., Kayne, P.S., and Grunstein, M. (1991). Yeast histone H4 N-terminal sequence is required for promoter activation in vivo. Cell *65*, 1023-1031.

Eissenberg, J.C., James, T.C., Foster-Hartnett, D.M., Hartnett, T., Ngan, V., and Elgin, S.C. (1990). Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. Proc Natl Acad Sci U S A *87*, 9923-9927.

Eissenberg, J.C., Morris, G.D., Reuter, G., and Hartnett, T. (1992). The heterochromatin-associated protein HP-1 is an essential protein in Drosophila with dosage-dependent effects on position-effect variegation. Genetics *131*, 345-352.

Eng, J., McCormack, A.L., and Yates III, J.R. (1994). Journal of the American Society of Mass Spectrometry *5*, 976-989.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

Falk, M.M., Grigera, P.R., Bergmann, I.E., Zibert, A., Multhaup, G., and Beck, E. (1990). Foot-and-mouth disease virus protease 3C induces specific proteolytic cleavage of host cell histone H3. J Virol *64*, 748-756.

Felbor, U., Kessler, B., Mothes, W., Goebel, H.H., Ploegh, H.L., Bronson, R.T., and Olsen, B.R. (2002). Neuronal loss and brain atrophy in mice lacking cathepsins B and L. Proc Natl Acad Sci U S A *99*, 7883-7888.

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

Fischle, W., Wang, Y., and Allis, C.D. (2003a). Binary switches and modification cassettes in histone biology and beyond. Nature *425*, 475-479.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003b). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev *17*, 1870-1881.

Francis, N.J., Follmer, N.E., Simon, M.D., Aghia, G., and Butler, J.D. (2009). Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. Cell *137*, 110-122.

Fujisaki, S., Ninomiya, Y., Ishihara, H., Miyazaki, M., Kanno, R., Asahara, T., and Kanno, M. (2003). Dimerization of the Polycomb-group protein Mel-18 is regulated by PKC phosphorylation. Biochem Biophys Res Commun *300*, 135-140.

Gan, Q., Yoshida, T., McDonald, O.G., and Owens, G.K. (2007). Concise review: epigenetic mechanisms contribute to pluripotency and cell lineage determination of embryonic stem cells. Stem Cells *25*, 2-9.

Garcia, B.A., Mollah, S., Ueberheide, B.M., Busby, S.A., Muratore, T.L., Shabanowitz, J., and Hunt, D.F. (2007). Chemical derivatization of histones for facilitated analysis by mass spectrometry. Nat Protoc *2*, 933-938.

Garcia, B.A., Thomas, C.E., Kelleher, N.L., and Mizzen, C.A. (2008). Tissue-specific expression and post-translational modification of histone H3 variants. J Proteome Res *7*, 4225-4236.

Gateff, E., Loffler, T., and Wismar, J. (1993). A temperature-sensitive brain tumor suppressor mutation of Drosophila melanogaster: developmental studies and molecular localization of the gene. Mech Dev *41*, 15-31.

Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L., Xu, M., Maynard, D.M., Yang, X., Shi, W., and Bryant, S.H. (2004). Open mass spectrometry search algorithm. J Proteome Res *3*, 958-964.

Georgel, P.T., Tsukiyama, T., and Wu, C. (1997). Role of histone tails in nucleosome remodeling by Drosophila NURF. EMBO J *16*, 4717-4726.

Giadrossi, S., Dvorkina, M., and Fisher, A.G. (2007). Chromatin organization and differentiation in embryonic stem cell models. Curr Opin Genet Dev *17*, 132-138.

Gil, J., Bernard, D., Martinez, D., and Beach, D. (2004). Polycomb CBX7 has a unifying role in cellular lifespan. Nat Cell Biol *6*, 67-72.

Goulet, B., Baruch, A., Moon, N.S., Poirier, M., Sansregret, L.L., Erickson, A., Bogyo, M., and Nepveu, A. (2004). A cathepsin L isoform that is devoid of a signal peptide localizes to the nucleus in S phase and processes the CDP/Cux transcription factor. Mol Cell *14*, 207-219.

Goulet, B., and Nepveu, A. (2004). Complete and limited proteolysis in cell cycle progression. Cell Cycle *3*, 986-989.

Grewal, S.I., and Elgin, S.C. (2002). Heterochromatin: new possibilities for the inheritance of structure. Curr Opin Genet Dev *12*, 178-187.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

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

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

Heard, E. (2004). Recent advances in X-chromosome inactivation. Curr Opin Cell Biol *16*, 247-255.

Heard, E., Chaumeil, J., Masui, O., and Okamoto, I. (2004). Mammalian X-chromosome inactivation: an epigenetics paradigm. Cold Spring Harb Symp Quant Biol *69*, 89-102.

Hiwasa, T., and Sakiyama, S. (1996). Nuclear localization of procathepsin L/MEP in ras-transformed mouse fibroblasts. Cancer Lett *99*, 87-91.

Hsu, J.Y., Sun, Z.W., Li, X., Reuben, M., Tatchell, K., Bishop, D.K., Grushcow, J.M., Brame, C.J., Caldwell, J.A., Hunt, D.F.*, et al.* (2000). Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. Cell *102*, 279-291.

Huang, R.C., and Bonner, J. (1962). Histone, a suppressor of chromosomal RNA synthesis. Proc Natl Acad Sci U S A *48*, 1216-1222.

Hudson, B.P., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2000). Solution structure and acetyl-lysine binding activity of the GCN5 bromodomain. J Mol Biol *304*, 355-370.

Irving, J.A., Shushanov, S.S., Pike, R.N., Popova, E.Y., Bromme, D., Coetzer, T.H., Bottomley, S.P., Boulynko, I.A., Grigoryev, S.A., and Whisstock, J.C. (2002). Inhibitory activity of a heterochromatin-associated serpin (MENT) against papain-like cysteine proteinases affects chromatin structure and blocks cell proliferation. J Biol Chem *277*, 13192-13201.

Ishidoh, K., Saido, T.C., Kawashima, S., Hirose, M., Watanabe, S., Sato, N., and Kominami, E. (1998). Multiple processing of procathepsin L to cathepsin L in vivo. Biochem Biophys Res Commun *252*, 202-207.

Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080-2083.

Jacobs, S.A., Taverna, S.D., Zhang, Y., Briggs, S.D., Li, J., Eissenberg, J.C., Allis, C.D., and Khorasanizadeh, S. (2001). Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. EMBO J *20*, 5232-5241.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422-1425.

Jin, C., and Felsenfeld, G. (2006). Distribution of histone H3.3 in hematopoietic cell lineages. Proc Natl Acad Sci U S A *103*, 574-579.

Johnson, L., Mollah, S., Garcia, B.A., Muratore, T.L., Shabanowitz, J., Hunt, D.F., and Jacobsen, S.E. (2004). Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Res *32*, 6511-6518.

Kayne, P.S., Kim, U.J., Han, M., Mullen, J.R., Yoshizaki, F., and Grunstein, M. (1988). Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. Cell *55*, 27-39.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049-1061.

Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., and Bedford, M.T. (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO Rep *7*, 397-403.

Klymenko, T., Papp, B., Fischle, W., Kocher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Muller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes Dev *20*, 1110-1122.

Kouzarides, T. (2002). Histone methylation in transcriptional control. Curr Opin Genet Dev *12*, 198-209.

Kurtin, P.J., Dewald, G.W., Shields, D.J., and Hanson, C.A. (1996). Hematologic disorders associated with deletions of chromosome 20q: a clinicopathologic study of 107 patients. Am J Clin Pathol *106*, 680-688.

Kuzmichev, A., Margueron, R., Vaquero, A., Preissner, T.S., Scher, M., Kirmizis, A., Ouyang, X., Brockdorff, N., Abate-Shen, C., Farnham, P.*, et al.* (2005). Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. Proc Natl Acad Sci U S A *102*, 1859-1864.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev *16*, 2893-2905.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Lan, F., Nottke, A.C., and Shi, Y. (2008). Mechanisms involved in the regulation of histone lysine demethylases. Curr Opin Cell Biol *20*, 316-325.

Landry, J., Sharov, A.A., Piao, Y., Sharova, L.V., Xiao, H., Southon, E., Matta, J., Tessarollo, L., Zhang, Y.E., Ko, M.S.*, et al.* (2008). Essential role of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells. PLoS Genet *4*, e1000241.

Lee, J.H., Hart, S.R., and Skalnik, D.G. (2004). Histone deacetylase activity is required for embryonic stem cell differentiation. Genesis *38*, 32-38.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.*, et al.* (2006a). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006b). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc *1*, 729-748.

Lennox, R.W., and Cohen, L.H. (1988). The production of tissue-specific histone complements during development. Biochem Cell Biol *66*, 636-649.

Li, H., Fischle, W., Wang, W., Duncan, E.M., Liang, L., Murakami-Ishibe, S., Allis, C.D., and Patel, D.J. (2007). Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. Mol Cell *28*, 677-691.

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

Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A., Ploegh, H., and Matsudaira, P. (2008). Molecular Cell Biology, 6th edn (W.H. Freeman).

Loyola, A., and Almouzni, G. (2007). Marking histone H3 variants: how, when and why? Trends Biochem Sci *32*, 425-433.

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

Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature *190*, 372-373.

Macdonald, N., Welburn, J.P., Noble, M.E., Nguyen, A., Yaffe, M.B., Clynes, D., Moggs, J.G., Orphanides, G., Thomson, S., Edmunds, J.W.*, et al.* (2005). Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14-3-3. Mol Cell *20*, 199-211.

MacGrogan, D., Alvarez, S., DeBlasio, T., Jhanwar, S.C., and Nimer, S.D. (2001). Identification of candidate genes on chromosome band 20q12 by physical mapping of translocation breakpoints found in myeloid leukemia cell lines. Oncogene *20*, 4150-4160.

Mann, R.K., and Grunstein, M. (1992). Histone H3 N-terminal mutations allow hyperactivation of the yeast GAL1 gene in vivo. EMBO J *11*, 3297-3306.

Marmorstein, R. (2001). Protein modules that manipulate histone tails for chromatin regulation. Nat Rev Mol Cell Biol *2*, 422-432.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A *78*, 7634-7638.

Martin, S.E., Shabanowitz, J., Hunt, D.F., and Marto, J.A. (2000). Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. Anal Chem *72*, 4266-4274.

Mason, R.W., Wilcox, D., Wikstrom, P., and Shaw, E.N. (1989). The identification of active forms of cysteine proteinases in Kirsten-virus-transformed mouse fibroblasts by use of a specific radiolabelled inhibitor. Biochem J *257*, 125-129.

Mendez, J., and Stillman, B. (2000). Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. Mol Cell Biol *20*, 8602-8612.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev Cell *10*, 105-116.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., and Hess, J.L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol Cell *10*, 1107-1117.

Milne, T.A., Dou, Y., Martin, M.E., Brock, H.W., Roeder, R.G., and Hess, J.L. (2005). MLL associates specifically with a subset of transcriptionally active target genes. Proc Natl Acad Sci U S A *102*, 14765-14770.

Min, J., Zhang, Y., and Xu, R.M. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. Genes Dev *17*, 1823-1828.

Mizuguchi, G., Tsukiyama, T., Wisniewski, J., and Wu, C. (1997). Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin. Mol Cell *1*, 141-150.

Morin, V., Sanchez, A., Quinones, K., Huidobro, J.G., Iribarren, C., Bustos, P., Puchi, M., Geneviere, A.M., and Imschenetzky, M. (2008). Cathepsin L inhibitor I blocks mitotic chromosomes decondensation during cleavage cell cycles of sea urchin embryos. J Cell Physiol.

Muchardt, C., Guilleme, M., Seeler, J.S., Trouche, D., Dejean, A., and Yaniv, M. (2002). Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha. EMBO Rep *3*, 975-981.

Muller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell *111*, 197-208.

Murry, C.E., and Keller, G. (2008). Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell *132*, 661-680.

Nakagawa, T., Roth, W., Wong, P., Nelson, A., Farr, A., Deussing, J., Villadangos, J.A., Ploegh, H., Peters, C., and Rudensky, A.Y. (1998). Cathepsin L: critical role in Ii degradation and CD4 T cell selection in the thymus. Science *280*, 450-453.

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

Ohno, S., Kaplan, W.D., and Kinosita, R. (1960). The basis of nuclear sex difference in somatic cells of the opossum Didelphis virginiana. Exp Cell Res *19*, 417-420.

Orlando, V., Jane, E.P., Chinwalla, V., Harte, P.J., and Paro, R. (1998). Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early Drosophila embryogenesis. EMBO J *17*, 5141-5150.

Otte, A.P., and Kwaks, T.H. (2003). Gene repression by Polycomb group protein complexes: a distinct complex for every occasion? Curr Opin Genet Dev *13*, 448-454.

Owen, D.J., Ornaghi, P., Yang, J.C., Lowe, N., Evans, P.R., Ballario, P., Neuhaus, D., Filetici, P., and Travers, A.A. (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. EMBO J *19*, 6141-6149.

Pajerowski, J.D., Dahl, K.N., Zhong, F.L., Sammak, P.J., and Discher, D.E. (2007). Physical plasticity of the nucleus in stem cell differentiation. Proc Natl Acad Sci U S A *104*, 15619-15624.

Park, S.H., and Raines, R.T. (2004). Fluorescence polarization assay to quantify protein-protein interactions. Methods Mol Biol *261*, 161-166.

Pearson, J.C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. Nat Rev Genet *6*, 893-904.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131-137.

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

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131-135.

Rando, O.J., and Ahmad, K. (2007). Rules and regulation in the primary structure of chromatin. Curr Opin Cell Biol *19*, 250-256.

Rawlings, N.D., and Barrett, A.J. MEROPS: the peptidase database (Cambridge CB10 1SA, UK).

Rawlings, N.D., Morton, F.R., Kok, C.Y., Kong, J., and Barrett, A.J. (2008). MEROPS: the peptidase database. Nucleic Acids Res *36*, D320-325.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D.*, et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Reijnen, M.J., Hamer, K.M., den Blaauwen, J.L., Lambrechts, C., Schoneveld, I., van Driel, R., and Otte, A.P. (1995). Polycomb and bmi-1 homologs are expressed in overlapping patterns in Xenopus embryos and are able to interact with each other. Mech Dev *53*, 35-46.

Riccio, M., Di Giaimo, R., Pianetti, S., Palmieri, P.P., Melli, M., and Santi, S. (2001). Nuclear localization of cystatin B, the cathepsin inhibitor implicated in myoclonus epilepsy (EPM1). Exp Cell Res *262*, 84-94.

Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu Rev Genet *38*, 413-443.

Roth, S.Y., Denu, J.M., and Allis, C.D. (2001). Histone acetyltransferases. Annu Rev Biochem *70*, 81-120.

Roth, W., Deussing, J., Botchkarev, V.A., Pauly-Evers, M., Saftig, P., Hafner, A., Schmidt, P., Schmahl, W., Scherer, J., Anton-Lamprecht, I.*, et al.* (2000). Cathepsin L deficiency as molecular defect of furless: hyperproliferation of keratinocytes and pertubation of hair follicle cycling. FASEB J *14*, 2075-2086.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol Cell *25*, 15-30.

Santos-Rosa, H., Kirmizis, A., Nelson, C., Bartke, T., Saksouk, N., Cote, J., and Kouzarides, T. (2009). Histone H3 tail clipping regulates gene expression. Nat Struct Mol Biol *16*, 17-22.

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

Schechter, I., and Berger, A. (1967). On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun *27*, 157-162.

Schneider, R., Bannister, A.J., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat Cell Biol *6*, 73-77.

Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. EMBO J *25*, 3110-3122.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735-745.

Shechter, D., Dormann, H.L., Allis, C.D., and Hake, S.B. (2007). Extraction, purification and analysis of histones. Nat Protoc *2*, 1445-1457.

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

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Simon, M.D., Chu, F., Racki, L.R., de la Cruz, C.C., Burlingame, A.L., Panning, B., Narlikar, G.J., and Shokat, K.M. (2007). The site-specific installation of methyl-lysine analogs into recombinant histones. Cell *128*, 1003-1012.

Simon, R.H., and Felsenfeld, G. (1979). A new procedure for purifying histone pairs H2A + H2B and H3 + H4 from chromatin using hydroxylapatite. Nucleic Acids Res *6*, 689-696.

Smith, A. (1991). Journal of Tissue Culture Methods *13*, 89-94.

Smith, A.G. (2001). Embryo-derived stem cells: of mice and men. Annu Rev Cell Dev Biol *17*, 435-462.

Sol-Church, K., Frenck, J., Bertenshaw, G., and Mason, R.W. (2000a). Characterization of mouse cathepsin R, a new member of a family of placentally expressed cysteine proteases. Biochim Biophys Acta *1492*, 488-492.

Sol-Church, K., Frenck, J., and Mason, R.W. (2000b). Cathepsin Q, a novel lysosomal cysteine protease highly expressed in placenta. Biochem Biophys Res Commun *267*, 791-795.

Sol-Church, K., Frenck, J., and Mason, R.W. (2000c). Mouse cathepsin M, a placenta-specific lysosomal cysteine protease related to cathepsins L and P. Biochim Biophys Acta *1491*, 289-294.

Stedman, E., and Stedman, E. (1951). The Basic Proteins of Cell Nuclei. Philosophical Transactions of the Royal Society, B *235*.

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

Sutcliffe, E.L., Parish, I.A., He, Y.Q., Juelich, T., Tierney, M.L., Rangasamy, D., Milburn, P.J., Parish, C.R., Tremethick, D.J., and Rao, S. (2009). Dynamic histone variant exchange accompanies gene induction in T cells. Mol Cell Biol *29*, 1972-1986.

Syka, J.E., Coon, J.J., Schroeder, M.J., Shabanowitz, J., and Hunt, D.F. (2004a). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A *101*, 9528-9533.

Syka, J.E., Marto, J.A., Bai, D.L., Horning, S., Senko, M.W., Schwartz, J.C., Ueberheide, B., Garcia, B., Busby, S., Muratore, T.*, et al.* (2004b). Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. J Proteome Res *3*, 621-626.

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

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science *272*, 408-411.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007a). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol *14*, 1025-1040.

Taverna, S.D., Ueberheide, B.M., Liu, Y., Tackett, A.J., Diaz, R.L., Shabanowitz, J., Chait, B.T., Hunt, D.F., and Allis, C.D. (2007b). Long-distance combinatorial linkage between methylation and acetylation on histone H3 N termini. Proc Natl Acad Sci U S A *104*, 2086-2091.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.

Torres-Padilla, M.E., Bannister, A.J., Hurd, P.J., Kouzarides, T., and Zernicka-Goetz, M. (2006). Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. Int J Dev Biol *50*, 455-461.

Trojer, P., Li, G., Sims, R.J., 3rd, Vaquero, A., Kalakonda, N., Boccuni, P., Lee, D., Erdjument-Bromage, H., Tempst, P., Nimer, S.D.*, et al.* (2007). L3MBTL1, a histone-methylation-dependent chromatin lock. Cell *129*, 915-928.

Tsukiyama, T., and Wu, C. (1995). Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell *83*, 1011-1020.

Turk, B., Turk, D., and Turk, V. (2000). Lysosomal cysteine proteases: more than scavengers. Biochim Biophys Acta *1477*, 98-111.

Van Holde, K.E. (1989). Chromatin (New York, Springer-Verlag).

Varga-Weisz, P.D., Blank, T.A., and Becker, P.B. (1995). Energy-dependent chromatin accessibility and nucleosome mobility in a cell-free system. EMBO J *14*, 2209-2216.

Vincenz, C., and Kerppola, T.K. (2008). Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. Proc Natl Acad Sci U S A *105*, 16572-16577.

Vogel, J.L., and Kristie, T.M. (2006). Site-specific proteolysis of the transcriptional coactivator HCF-1 can regulate its interaction with protein cofactors. Proc Natl Acad Sci U S A *103*, 6817-6822.

Wang, P.W., Eisenbart, J.D., Espinosa, R., 3rd, Davis, E.M., Larson, R.A., and Le Beau, M.M. (2000). Refinement of the smallest commonly deleted segment of chromosome 20 in malignant myeloid diseases and development of a PAC-based physical and transcription map. Genomics *67*, 28-39.

Wang, W.K., Tereshko, V., Boccuni, P., MacGrogan, D., Nimer, S.D., and Patel, D.J. (2003). Malignant brain tumor repeats: a three-leaved propeller architecture with ligand/peptide binding pockets. Structure *11*, 775-789.

Wells, D., Hoffman, D., and Kedes, L. (1987). Unusual structure, evolutionary conservation of non-coding sequences and numerous pseudogenes characterize the human H3.3 histone multigene family. Nucleic Acids Res *15*, 2871-2889.

Whitcomb, S.J., Basu, A., Allis, C.D., and Bernstein, E. (2007). Polycomb Group proteins: an evolutionary perspective. Trends Genet *23*, 494-502.

Winslow, T., and Duckwall, C. (2001).

Wismar, J., Loffler, T., Habtemichael, N., Vef, O., Geissen, M., Zirwes, R., Altmeyer, W., Sass, H., and Gateff, E. (1995). The Drosophila melanogaster tumor suppressor gene lethal(3)malignant brain tumor encodes a proline-rich protein with a novel zinc finger. Mech Dev *53*, 141-154.

Wright, W.W., Smith, L., Kerr, C., and Charron, M. (2003). Mice that express enzymatically inactive cathepsin L exhibit abnormal spermatogenesis. Biol Reprod *68*, 680-687.

Wysocka, J. (2006). Identifying novel proteins recognizing histone modifications using peptide pull-down assay. Methods *40*, 339-343.

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

Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H., and Allis, C.D. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell *121*, 859-872.

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

Yu, L., and Gorovsky, M.A. (1997). Constitutive expression, not a particular primary sequence, is the important feature of the H3 replacement variant hv2 in Tetrahymena thermophila. Mol Cell Biol *17*, 6303-6310.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750-756.