THE HISTONE H3 LYSINE 36 METHYLTRANSFERASE HYPB FACILITATES EMBRYONIC STEM CELL DIFFERENTIATION

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

Von

Tianke Wang

Aus China

Basel, 2007

Friedrich Miescher Institute for Biomedical Research

Maulbeerstrasse 66, CH-4058, Basel

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

Auf Antrag von

Prof. Susan Gasser, Prof. Primo Schär

Basel, den 16 October 2007

Prof. Dr. Hans-Peter Hauri

SUMMARY

Histone H3 lysine 36 (H3K36) methylation was identified as a conserved modification from yeast to human. In yeast, biochemical characterization of the SET2 protein and genome wide mapping of H3K36me2 and K36me3 indicate that H3K36 methylation functions in transcription elongation through Set2/Rpd3S pathway.

A number of H3K36 methyltransferases and demethylases have been identified in different species, which underscores the dynamics of H3K36 methylation. As in yeast, H3K36me3 also peaks at the 3' end of genes in mammals. The genome wide view of H3K36me1 and H3K36me2 is not clear yet. To date, the functional significance of H3K36 methylation remains largely unknown in mammals.

In this thesis, homologs of SET2 in mammals, including Nsd1, Nsd2, Nsd3, and HypB, were studied. Nsd proteins displayed weak methyltransferase activity towards histone H3 *in vitro*. Their target specificities needs to be further analyzed. *In vitro*, HypB showed strong activity for histone H3 lysine 36. *In vivo*, H3K36 trimethylation levels were significantly reduced in HypB knock-down cells, indicating that HypB is a major H3K36 trimethyltransferase. Distribution of H3K36 methylation (mono-, di-, and tri-) were analyzed by immunofluorescence both in human and mouse cells. All three methylation states of showed euchromatic distribution, whereas H3K36 mono- and dimethylation also showed heterochromatic enrichment in terminally differentiated NIH3T3 cells as well. In embryonic stem cells, H3K36 methylation showed an inverse correlation with the expression level of Oct4, a stem cell marker, suggesting a potential role of H3K36 methylation in ES cell differentiation. After induction of differentiation by removing LIF or adding retinoic acid to the culture medium, stem cell genes failed to be repressed and lineage specific genes failed to be activated to the same degree in HypB knockdown cell as observed in mock treated ES cells. The presence of H3K36me3 along *Oct4* locus was mapped by CHIP. H3K36me3 was highly enriched in the coding region, and was low upstream of the transcription start site in undifferentiated ES cells. During differentiation, however, H3K36me3 decreased on the coding region and increased slightly on enhancer region of *Oct4* in the course of Oct4 repression after differentiation. In all, we propose that H3K36me3 is catalyzed by HypB and has an inverse correlation with Oct4 expression. HypB facilitates ES cell differentiation. The molecular mechanism by which HypB facilitates differentiation requires further investigation.

TABLE OF CONTENTS

Summary..... 1

Table of contents.....2

List of figures..... 4

Abbreviation 7

Chapter 1. Introduction.....10

1.1 Genome and epigenome.....10

1.2 Higher-order chromatin organization.....11

1.3 Euchromatin and heterochromatin.....12

1.4 Epigenetic regulation of chromatin function..... ...13

1.4.1 Chromatin modifications.....13

1.4.1.1 Histone modifications and histone-modifying enzymes.....13

1.4.1.2 Mechanisms of histone modifications.....17

1.4.1.3 Functional consequence of histone modifications.....18

1.4.2 Chromatin remodelling and histone eviction.....18

1.4.3 Histone variants incorporation.....18

1.5 Histone acetylation and deacetylation.....18

1.6 Histone methylation.....20

1.7 DNA methylation.....26

1.8 H3K36 methylation.....28

1.9 Genetic and epigenetic regulators of pluripotency.....35

1.9.1 ES cell self renewal and differentiation.....35

1.9.2 Genetic regulators of pluripotency.....35

1.9.3 Chromatin regulators of pluripotency.....38

Chapter 2. Results.....43

2.1 Distribution of H3K36 methylation in mammalian cells.....43

2.1.1 Distribution of H3K36 methylation in differentiated mammalian cells.....43

2.1.2 Distribution of H3K36 methylation in undifferentiated embryonic stem cells...50

2.2 Identification of H3K36 methyltransferase *in vitro*.....53

2.2.1 *In vitro* HMT assay of Nsd proteins.....54

2.2.2 *In vitro* HMT assay of HypB.....57

2.3 Identification of H3K36 methyltransferase *in vivo*.....59

2.4 HypB facilitates ES cell differentiation.....68

2.4.1 The effect of HypB knockdown in undifferentiated cells……..**68**
2.4.2 The effect of HypB knockdown in differentiated cells ……**71**
2.5 Dynamics of H3K36me3 and Oct4 expression during differentiation……**75**
2.6 Gene expression analysis in HypB knockdown ES cells……**77**
2.7 Distribution of H3K36me3 on *Oct4* locus during differentiation……**84**
Chapter 3. Discussion……88
3.1 HypB is the major H3K36me3 methyltransferase in mammalian cells……**88**
3.2 HypB facilitates ES cell differentiation: a role within the coding region of target gene?……**90**
3.3 HypB facilitates ES cell differentiation: a role at promoter and enhancer of *Oct4*?..**91**
3.4 The dynamic of H3K36 methylation on chromatin……..**93**
3.5 Additional roles of H3K36 methylation……...**98**
3.6 H3K36 methyltransferase and cancers……..**99**
3.7 Conclusion……**100**
Chapter 4. Materials and methods……101
4.1 Materials……....**101**
4.2 Methods……..**107**
Chapter 5. Appendix……113
References……...116
Acknowledgement……129
Curriculum Vitae……130
Erklärung……..133

LIST OF FIGURES

Chapter 1

Figure 1. DNA versus Chromatin
Figure 2. Organization of DNA within chromatin structure
Figure 3. Distinction between euchromatin and heterochromatin domains
Figure 4. Histone modifying enzymes
Figure 5. Genome wide distribution pattern of histone modifications from a transcription perspective
Figure 6. Models showing how histone modifications affect chromatin template
Figure 7. Recruitment of proteins to histones
Figure 8. Structure feature of SET2 family HMTs
Figure 9. Function of H3K36 methylation in transcription
Figure 10. Mechanism of H3K36 methylation by SET2 during transcriptional elongation
Figure 11. A transcription factor network to control ES cell self-renewal and differentiation
Figure 12. Bivalent domains in ES cells
Figure 13. Model for *Oct4* heterochromatinization
Figure 14. Model of *Oct4* gene repression and silencing initiated by GCNF dependent recruitment of Mbd2 and Mbd3

Chapter 2

Figure 15. Distribution pattern of H3K36 methylation in mouse NIH3T3 cells
Figure 16. Distribution pattern of H3K36 methylation in human 293 cells
Figure 17. Distribution pattern of H3K36 methylation in U2OS cells
Figure 18. FACS sorting profile of NIH3T3 cells
Figure 19. Distribution pattern of H3K36me1 during cell cycle in NIH3T3 cells
Figure 20. Distribution pattern of H3K36me2 during cell cycle in NIH3T3 cells
Figure 21. Distribution pattern of H3K36me3 during cell cycle in NIH3T3 cells
Figure 22. Inverse correlation between H3K36 methylation and Oct4 expression
Figure 23. H3K36me1 and H3K36me2 showed heterochromatic foci in cells with low level of Oct4
Figure 24. Inverse correlation between H3K36me3 and Oct4 expression throughout cell cycle

Figure 25. Inverse correlation between H3K36me3 and Oct4 expression in PGK12 and LF2 ES cells
Figure 26. A protein sequence alignment of SET domains of SET2 family HMTs
Figure 27. Expression of recombinant GST fusion proteins
Figure 28. *In vitro* HMT assay of Nsd proteins
Figure 29. HypB methylated histone H3 *in vitro*
Figure 30. HypB methylated histone H3 peptide at lysine 36
Figure 31. Specificity of HypB on nucleosomes substrates
Figure 32. Gene expression in different human and mouse cell lines
Figure 33. Map of shRNA expressing vector
Figure 34. Knockdown of NSD1, NSD2, NSD3, and HYPB in 293 cells
Figure 35. H3K36me3 was strongly reduced upon HYPB knockdown by shRNA in 293 cells detected by IF
Figure 36. H3K36me3 was strongly reduced upon HYPB knockdown by shRNA in 293 cells detected by western blot
Figure 37. H3K36me3 was strongly reduced upon HYPB knockdown by shRNA in U2OS cells detected by IF
Figure 38. H3K36me3 was strongly reduce upon HYPB knockdown by stealth RNA in U2OS cells detected by IF
Figure 39. H3K36 was strongly reduced upon HypB knockdown by stealth RNA in NIH3T3 and CCE cells
Figure 40. No apparent change of H3K36me1 and H3K36me2 was observed upon HypB knockdown by stealth RNA in CCE cells
Figure 41. HypB knockdown CCE cells proliferated faster than control cells
Figure 42. Cell cycle analysis of HypB knockdown CCE ES cells
Figure 43. IF of undifferentiated HypB knockdown and control CCE ES cells
Figure 44. Knockdown of HypB in CCE cells
Figure 45. Oct4 repression defects in HypB knockdown ES cells under promotion of differentiation
Figure 46. H3K36me3 in mock treated and HypB knockdown cells
Figure 47. Dynamic of HypB, Oct4, and H3K36me3 during differentiation
Figure 48. HypB knockdown ES cells fail to repress ES cell markers as control cells upon induction of differentiation

Figure 49. HypB knockdown ES cells fail to activate differentiation-specific genes as control cells upon induction of differentiation
Figure 50. Kinetics of Gata6 and Nestin during differentiation
Figure 51. Dppa3 was downregulated in HypB knockdown ES cells
Figure 52. Stable knockdown of HypB in CCE cells
Figure 53. Scheme picture of *Oct4* locus and primers
Figure 54. Distribution of H3K36me3 on the *Oct4* locus in undifferentiated and differentiating ES cells
Figure 55. Distribution of H3K4me3 on the *Oct4* locus in undifferentiated and differentiating ES cells

Chapter 3

Figure 56. Working model of H3K36me3 on the promoter of *Oct4* during differentiation
Figure 57. Localization of H3K36me2 in mouse oocyte and early embryo
Figure 58. Localization of overexpressed Nsd2, Nsd3, and HypB
Figure 59. Localization of H3K36me3 in mouse oocyte and early embryo
Figure 60. H3K36me3 was excluded from DNA double stands break
Figure 61. Downregulation of HYPB in leukemia

ABBREVIATIONS

µM	micromolar
7-AAD	7-amino-actinomycin D
amp	ampicilline
ATP	adenosine triphosphate
BMP	bone morphogenic protein
bp	base pair
BrdU	bromodeoxyuridine
CBP	C/EBP binding protein
cDNA	complementary DNA
Ce	*Caenorhabditis elegans*
CHD3	chromodomain helicase DNA binding protein 3
ChIP	Chromatin immunoprecipitation
CTD	carboxy terminal domain
DAPI	4, 6-Diamidino-2-phenylindole
DE	Distal enhancer
Dm	*Drosophila melanogaster*
DMEM	Dulbecco's modified Eagles medium
DNA	desoxyribonucleic acid
dNTP	desoxynucleoside triphosphate
DTT	dithiothreitol
Eaf3	Esa1p-Associated Factor 3
EDTA	ethylenediamintetraacetate
ES	embryonic stem
FACS	Fluorescence activated cell sorting
FACT	facilitates chromatin transcription
FCS	fetal calf serum
GCNF	germ cell nuclear factor
GST	glutathione S-transferase
H1	histone 1
H2A/B	histone 2A/B
H3	histone 3
H4	histone 4
HAT	histone acetyl transferase
HDAC	histone deacetylase
HMT	histone methyl transferase
HP1	heterochromati protein 1
HypB	huntingtin yeast partner B

IF	Immunofluorescence
IP	immunoprecipitation
IPTG	isopropyl-β-thiogalactoside
kb	kilobase-pair
kD	kiloDalton
LIF	Leukemia Inhibitory Factor
M	mole
Mbd2	methyl-CpG binding domain protein 2
Mbd3	methyl-CpG binding domain protein 3
MCS	multiple cloning site
Mm	*Mus musculus*
mol	molar
Mw	molecular weight
N-CoR	nuclear hormone receptor co-repressor
nM	nanomolar
NP40	Nonidet P 40
Nsd1	Nuclear receptor-binding, SET domain-containing protein 1
Nsd2	Nuclear SET domain-containing protein 2
Nsd3	Nuclear SET domain-containing protein 3
NuRD	nucleosome remodelling and deacetylation complex
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	polymerase chain reaction
PE	proximal enhancer
PP	proximal promoter
PVDF	polyvinyldifluoride
RA	retinoic acid
RNA	ribonucleic acid
RNAi	ribonucleic acid interference
Rpd3	Reduced Potassium Dependency 3
Rpd3S	small Rpd3 complex
Sc	*Saccharomyces cerevisiae*
SDS	sodium dodecylsulfate
SET	Su(var)3-9, Enhancer-of-zeste, Trithorax
shRNA	short hairpin RNA
siRNA	short interfering RNA
SMART	Simple Modular Architecture Research Tool
SRI	SET2 Rbp1 Interacting
SUMO	small ubiquitin like modifier

SWI/SNF	switch from sugar non fermenting-chromatin remodeling machine
TBP	TATA box binding protein
TF	transcription factor
Tris	Tris (hydroxymethyl)-aminoethan
TS	trophoblast stem
UAS	upstream activating sequences
Wnt	wingless-type MMTV integration site family

CHAPTER 1 INTRODUCTION

1.1 Genome and epigenome

Genomic DNA of eukaryotic cells is highly folded and compacted with histone and non-histone proteins into chromatin. Gene expression, chromosome segregation, DNA replication, repair, and recombination all act not on DNA alone, but on the chromatin template. Chemical modifications occur on DNA and histone proteins and form a complex regulatory network that modulates chromatin structure and genome function. The epigenome refers to the complete description of these potentially heritable changes across the genome (Bernstein et al., 2007). Thus, a single genome can generate a multitude of distinct 'epigenomes', as the fertilized egg progresses through development and translates its information into more than 200 different types of cell fates (Allis et al., 2006) (Figure 1).

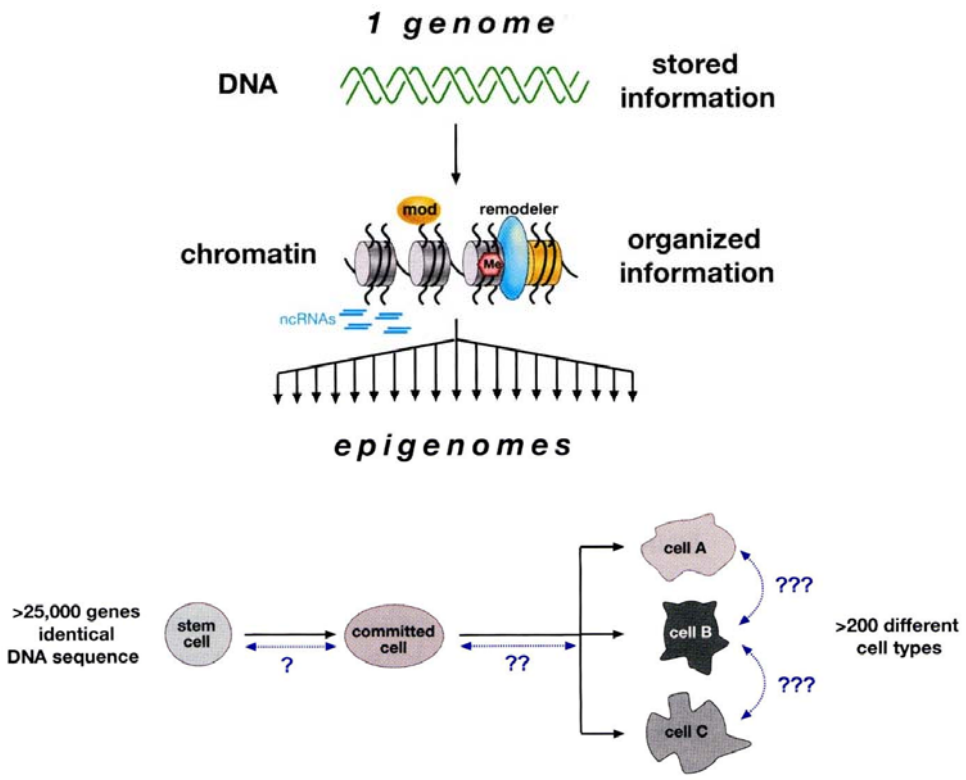

Figure 1. DNA Versus Chromatin

The genome: Invariant DNA sequence (green double helix) if an individual. The epigenome: the overall chromatin composition, which indexes the entire genome in any given cell. It varies according to cell type and response to internal and external signals it receives. (Lower panel) Epigenome diversification occurs during development in multicellular organisms as differentiation proceeds from a single stem cell (the fertilized embryo) to more committed cells. Reversal of differentiation or transdifferentiation (blue lines) requires the reprogramming of a cell's epigenome. (Allis et al. 2006, Epigenetics)

1.2 Higher-order chromatin organization

The fundamental subunit of chromatin is the nucleosome, which consists of approximately 147 base pairs (bp) of DNA wrapped around an octamer of the four core histones (H3, H4, H2A, H2B) (Luger et al., 1997) (Felsenfeld and Groudine, 2003). Specific interactions between individual nucleosomes drive the folding of a nucleosomal array (the primary structure of chromatin) into the 30 nm fiber (secondary structure) and into larger-scale configurations (tertiary structures) that build an entire chromosome (Tremethick, 2007) (Figure 2).

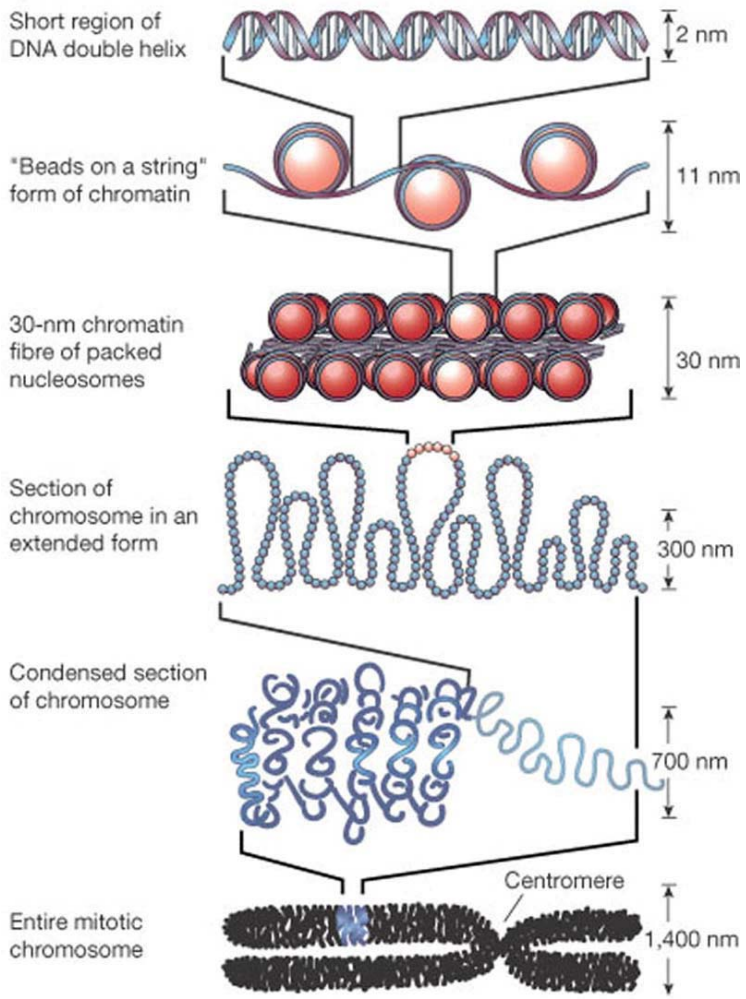

Figure 2. Organization of DNA within the chromatin structure

The lowest level of organization is the nucleosome, in which two superhelical turns of DNA (a total of 147 base pairs) are wound around the outside of a histone octamer. Nucleosomes are connected to one another by short stretches of linker DNA. At the next level of organization the string of nucleosomes is folded into a fibre about 30 nm in diameter, and these fibres are then further folded into higher-order structures. At levels of structure beyond the nucleosome the details of folding are still uncertain. (Felsenfeld and Groudine, 2003)

1.3 Euchromatin and heterochromatin

In a simple classification approach, there are two different types of chromatin environments in the genome, active euchromatin and silent heterochromatin. Euchromatin exists in an “open” (decompacted), more nuclease-sensitive configuration, making it “poised” for gene expression, although not necessarily being transcriptionally active. Euchromatin consists largely of coding sequences, which only account for a small fraction (4%) of the genome in mammals. Heterochromatin exists in a “closed” (compacted) configuration, making it transcriptionally limited. Heterochromatin mainly consists of noncoding and/or repetitive sequences. Heterochromatin plays a critically important role in the organization and proper function of genomes (Figure 3).

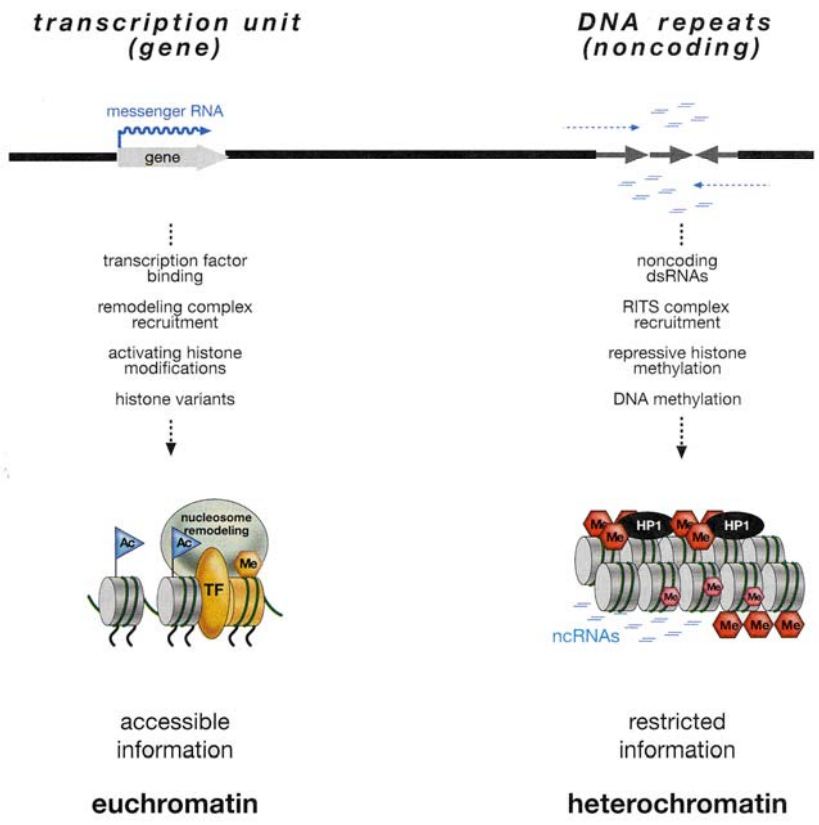

Figure 3. Distinction between Euchromatin and Heterochromatin Momains

Summary of common differences between euchromatin and constitutive heterochromatin. This includes in the type of transcripts produced, recruitment of DNA-binding proteins (i.e. transcription factor [TF]), chromatin-associated proteins and complexes, covalent histone modifications, and histone variant composition. (Allis et al. 2006, Epigenetics)

1.4 Epigenetic regulation of chromatin function

The dynamics of chromatin structure is tightly regulated through multiple mechanisms including chromatin modification, chromatin remodeling, histone variant incorporation and histone eviction (Li et al., 2007a).

1.4.1 Chromatin modifications

Chromatin modifications fall into two main categories: histone modifications and DNA methylation (Bernstein et al., 2007) (Berger, 2007) (Table 1). Histone modification has been found to be a central feature of genomic regulation.

1.4.1.1 Histone modifications and histone-modifying enzymes

Both histone tails and globular domains are subject to a vast array of posttranslational modifications. These modifications include methylation of arginine (R) residues; methylation, acetylation, ubiqutylation, ADP-ribosylation, and sumoylation of lysine (K) residues; and phosphorylation of serines and threonines (Table 1). Enzymes that direct modifications have been identified over the past ten years (Table 2). Most modifications have been found to be dynamic and enzymes that remove the modification have been identified. Enzymes that catalyse histone modifications and their counterpart enzymes that reverse the modifications antagonistically, govern the steady-state balance of each modification (Allis et al., 2006) (Figure 4).

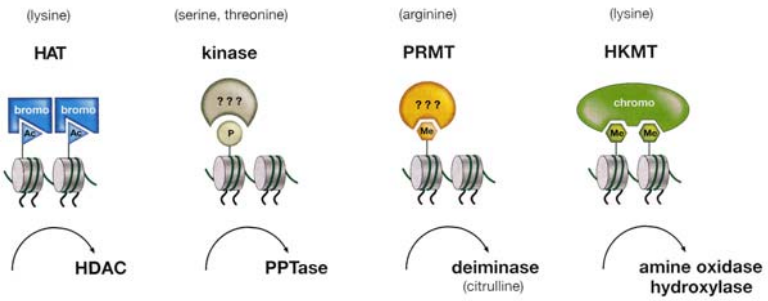

Figure 4. Histone Modifying Enzymes

Covalent histone modifications are transduced by histone-modifying enzymes ("writer") abd removed by antagonizing activities. They are classified into families according to the type of enzymatic action (e.g. acetylation or phosphorylation). Protein domains with specific affinity for a histone tail modification are termed "readers". (HAT) Histone acetyltransferase; (PRMT) protein argine methyltransferase; (HKMT) histone lysine methyltransferase; (HDAC) histone deacetylase; (PPTase) protein phosphatase: (Ac) acetylation; (P) phosphorylation; (Me) methylation. (Allis et al. 2006, Epigenetics)

Histone acetyltransferases (HATs) acetylate specific lysine residues in histone substrates and are reversed by the action of histone deacetylases (HDACs). The histone kinase family of enzymes phosphorylate specific serine or threonine residues, and phosphatases (PPTase) remove phosphorylation marks. Protein arginine methyltransferases (PRMTs) methylate arginine residues, and are indirectly reversed by the action of deiminases, which convert methyl-arginine (or arginine) to a citrulline residue. Histone lysine methyltransferases (HKMTs) methylate lysine residue of histone, and this mark is erased by histone demethylase.

Table1. Chromatin modifications

Mark*	Transcriptionally relevant sites	Transcriptional role
DNA methylation		
Methylated cytosine (meC)	CpG islands	Repression
Histone modifications		
Acetylated lysine (Kac)	H3 (K9,K14,K18,K36,K56)	Activation
	H4 (K5,K8,K12,K16)	Activation
	H2A	Activation
	H2B (K6,K7,K16,K17)	Activation
Phosphorylated serine/threonine (S/Tph)	H3 (T3,S10,S28)	Activation
	H2A	Activation
	H2B	Activation
Methylated arginine (Rme)	H3 (R2,R17,R26)	Activation
	H4 (R3)	Activation
Methylated lysine (Kme)	H3 (K4,K36,K79)	Activation
	H3 (K9,K27)	Repression
	K4 (K20)	Repression
Ubiquitylated lysine (Kub)	H2B (K123§/120¶)	Activation
	H2A (K119¶)	Repression
Sumoylation lysine (Ksu)	H4 (K5,K8,K12,K16)	Repression
	H2A (K126)	Repression
	H2B (K6,K7,K16,K17)	Repression
Isomerized proline (Pisom)	H3 (P30,P38)	Activation/Repression

* The modification on either DNA or a histone.

§ Yeast (*Saccharomyces cerevisiae*).

¶ Mammals

(Adapted from Berger, 2007)

Table2. Chromatin-modifying enzymes

Modifications	Position		*S. cerevisiae*	*S. pombe*	*Drosophila*	Mammals
Methylation	H3	K4	Set1	Set1	Trx, Ash1	Mll, All-1, Ash1, ALR-1/2, ALR, Set7/9, Set1
		K9	n/a	Clr4	Su(var)3-9 Ash1	Suv39h, G9a, SETDB1, CLL8, Eu-HMTase1
		K27			E(Z)	Ezh2
		K36	Set2	Set2	dSet2	HypB, Smyd2, NSD1, Ash1
		K79	Dot1			Dot1L
	H4	K20		Set9	PR-Set7, Ash1	PR-Set7, Set8
Arg methylation	H3	R2				CARM1
		R17				CARM1
		R26				CARM1
	H4	R3				PRMT1
Phosphorylation	H3	S10	Snf1			
		S28				MSK1, MSK2
Ubiquitylation	H2A	K119				Bmi/Ring1a
	H2B	K120/123	Rad6, Bre1	Rad6		BbcH6, RNF20/40
Isomerization	H3	P10, P38	FPR4			
Acetylation	H3	K9				PCAF/GCN5
		K14	Sas3			CBP/P300, TIP60, PCAF/GCN5
		K18				CBP/P300, PCAF/GCN5
		K56	Rtt109			
	H4	K5	Esa1	Mst1		HAT1,CBP/P300, TIP60, HBO1
		K8	Esa1	Mst1		CBP/P300, TIP60, HBO1
		K12	Esa1	Mst1		HAT1, TIP60, HBO1
		K16	Sas2, NuA4	Mst2	dMOF	TIP60, hMOF
	H2A	K5				CBP/P300
	H2B	K12				CBP/P300
		K15				CBP/P300

(Adapted from Li et al. 2007 and Kouzarides, 2007)

Most modifications have distinct localization patterns associated with different chromatin environments. Generally, histone acetylation, methylation at H3K4 and H3K79 has been linked to activation of transcription and referred to as euchromatin modifications. Methylation at H3K9, H3K27 and H4K20 has been linked to repression and termed as heterochromatin modifications. Methylation at H3K36 has been linked to transcriptional elongation. However, it is involved in a deacetylation pathway to restore a repressive chromatin after passage of RNA polymerase II (see also 1.8). Within a transcription unit, most modifications are distributed in distinct localized patterns (Figure 5). The localization of histone modification is tightly regulated and is crucial for its effect on chromatin structure and transcription (Li et al., 2007a).

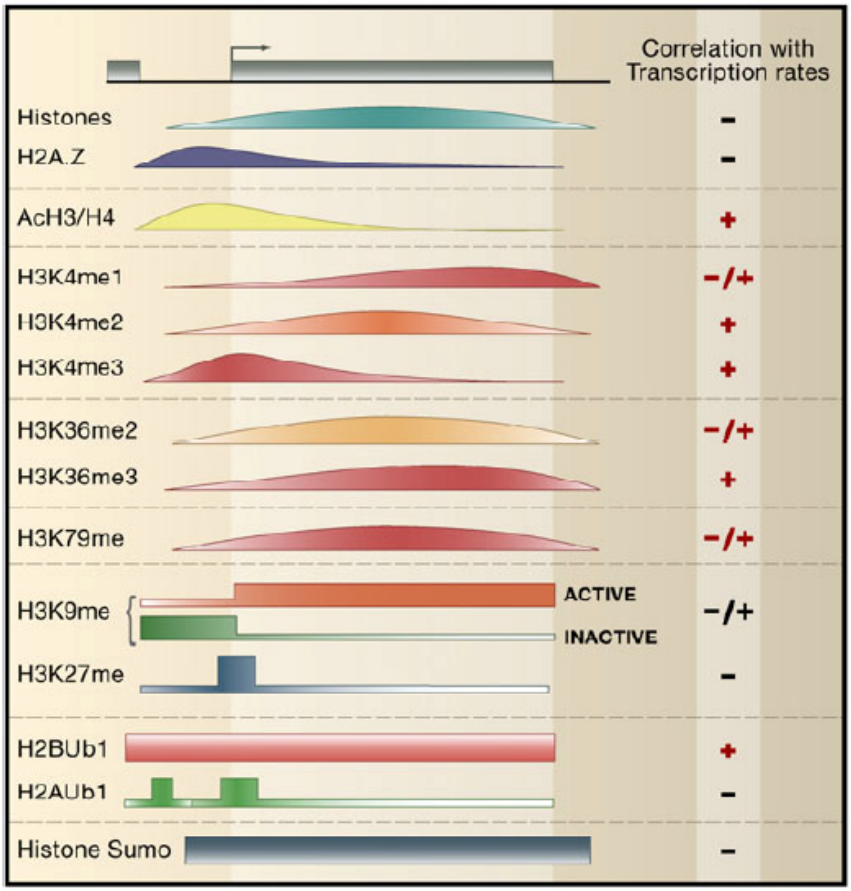

Figure 5. Genome-Wide Distribution Pattern of Histone Modifications from a Transcription Perspective

The distribution of histones and their modifications are mapped on an arbitrary gene relative to its promoter the upstream region, the core promoter, ORF, 5' end of ORF and 3' end of ORF. The curves represent the patterns that are determined via genome-wide approaches. The squares indicate that the data are based on only a few case studies. With the exception of the data on K9 and K27 methylation, most of the data are based on yeast genes. (Li et al, 2007)

1.4.1.2 Functional mechanisms of histone modifications

There are three characterized mechanisms for the function of modifications (Figure 6). First, histone modifications may affect the chromatin structure by changing the contact between different histones in adjacent nucleosomes or the interaction of histone with DNA, which are brought about by changes in the physical properties of modified histone tails, such as a modulation in the electrostatic charge or tail structure. This mechanism is considered as a *cis-effect*. Acetylation and phosphorylation are thought to function through this mechanism. Secondly, histone modifications may prevent or disrupt the binding of proteins that associate with chromatin or histones. Thirdly, histone modifications may provide binding sites that attract certain effectors via specific domains. Methylation is recognized by chromo-like domains of the Royal family (chromo, tudor, MBT) and PHD domains, acetylation is recognized by bromodomains, and phosphorylation is recognized by a domain present in 14-3-3 proteins. The latter two mechanisms are considered as *trans-effects* (Kouzarides, 2007) (Figure 7).

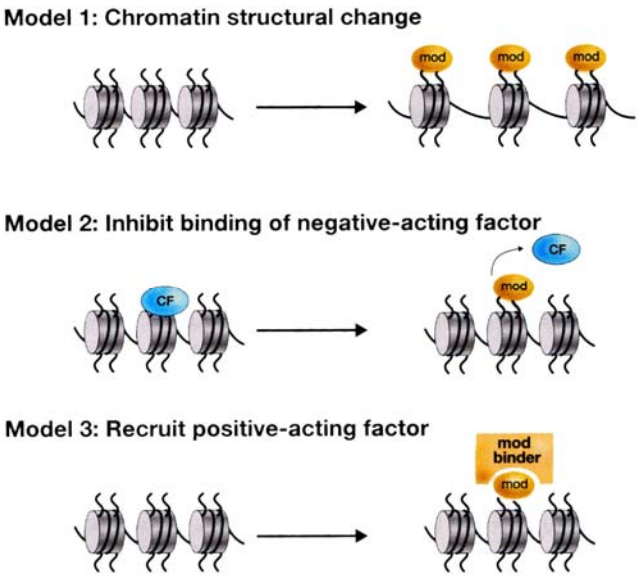

Figure 6. Models showing how Histones Posttranslational Modifications Affect the Chromatin Template

Model 1 propose that changes to chromatin structure are mediated by the cis effects of covalent histone modifications, such as histone acetylation or phosphorylation. Model 2 illustrates the inhibitory effect of a histone modification for the binding of a chromatin-associated factor (CF), as exemplified by H3S10 phosphorylation occluding HP1 binding at methylated H3K9. In model 3, a histone modification may provide binding specificity for a chromatin-associated factor. A classic example is HP1 binding through its chromodomain to methylated H3K9. (Allis et al. 2006, Epigenetics)

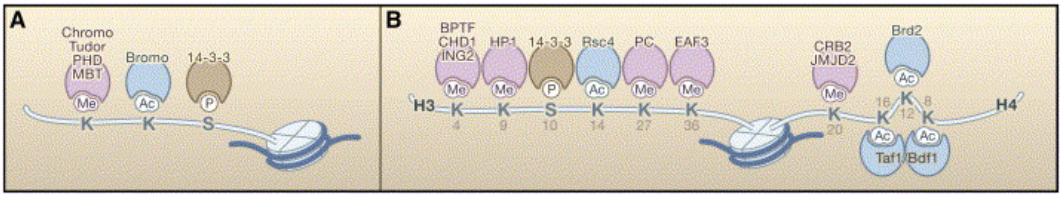

Figure 7. Recruitment of Proteins to Histones.

(A) Domains used for the recognition of methylated lysines, acetylated lysines, or phosphorylated serines. (B) Proteins found that associate preferentially with modified versions of histone H3 and histone H4. (Kouzarides, 2007)

Furthermore, crosstalk between different modifications is also implicated in the function of histone modification at different levels. Different modifications on the same residue antagonize each other. Certain modification of one residue may affect the modification or the affinity binding of effector proteins of adjacent resides within the same histone tail (Mateescu et al., 2004). Modification of one residue may also affect the modifications of the other residues on different histone tails.

1.4.1.3 Functional consequence of histone modifications

The function of histone modifications includes the establishment of global chromatin environments (such as euchromatin and heterochromatin) and the orchestration of DNA-based biological process such as transcription, replication, DNA repair, and chromosome condensation. One hypothesis is that there is a “Histone code”, linking specific modifications with individual processes (Strahl and Allis, 2000; Turner, 2000). However, It is not likely that the histone code functions as a genetic code that is a predictable, invariant and nearly universal language. The function of different histone modifications varies considerably from one organism to the other, especially between lower and higher eukaryotes. One limitation of the histone code is that one modification does not invariantly translate to one biological output. Thus, a more general hypothesis has been proposed where histone modification serve as a nuclear DNA-associated signal transduction pathway (Allis et al., 2006).

1.4.2 Chromatin remodeling and histone eviction

The second major class of chromatin regulators are protein complexes that utilize ATP hydrolysis to slide nucleosomes, replace histones, or alter the histone-DNA contacts. Due to this, they are generally referred to as chromatin-remodeling complexes. The consequences of remodeling include transient unwrapping of DNA from histone octamers, forming DNA loops, or moving nucleosomes to different translational positions, all of which change the accessibility of nucleosomal DNA to transcription factors (Flaus and Owen-Hughes, 2004; Saha et al., 2006). Recent studies reveal that histone displacement does occur *in vivo* during chromatin remodeling and that entire histone octamers can also be displaced (evicted) or exchanged under certain circumstances (Li et al., 2007a).

1.4.3 Histone variants incorporation

It is known that many variant forms of histones exist in many different organisms. Variants have been found to distinguish alternative chromatin states at centromeres, at the inactive mammalian X chromosome and at transcriptionally active loci, and are assembled into chromatin by specific protein complexes. The differences between variants and canonical histones in the histone tails, in the histone fold domains, or the difference of key amino acid residues may result in different nucleosome structures, stabilities, modifications, and compositions. The incorporation of histone variants impacts higher-order chromatin structure in various ways to facilitate various cellular processes (Henikoff et al., 2004; Kamakaka and Biggins, 2005; Li et al., 2007a).

1.5 Histone acetylation and deacetylation

Most of the acetylation sites fall within the N-terminal tail of the histones. However, a lysine within the core domain of H3K56 has recently been found to be acetylated (Xu et al., 2005) (Table 1). Histone acetylation is enriched predominantly at the promoter region and the transcriptional start site of active genes, and this enrichment drops substantially across the ORFs (Pokholok et al., 2005).

Acetyltransferases are divided into three main families: GANT, MYST, and CBP/p300 (Sterner and Berger, 2000). In general, these enzymes modify more than one lysine but some limited specificity can be detected for some enzymes (Table 2). GANT targets

histone H3 as its main substrate, the MYST family targets histone H4 as its main substrate and CBP/p300 targets both H3 and H4.

The reversal of acetylation correlates with transcriptional repression. There are three distinct families of histone deacetylases: the class I and class II histone deacetylase and the class III NAD-dependant enzymes of the Sir family (Ouaissi and Ouaissi, 2006). These enzymes are involved in multiple signalling pathways and they are present in numerous repressive chromatin complexes. In general, these enzymes do not show specificity for a particular acetyl group, although some of the yeast enzymes have specificity for a particular histone: Hda1 for H3 and H2B (Wu et al., 2001); Hos2 for H3 and H4 (Wang et al., 2002a). The fission yeast class III deacetylase Sir2 has some selectivity for H4K16ac, and recently, the human Sir family member SirT2 has been demonstrated to have a similar preference (Vaquero et al., 2006).

Bromodomains were discovered as acetyl-lysine binding domains. It is widely distributed among the different chromatin modifying proteins, such as histone acetyltransferase (GCNF/PANF, CBP/p300, $TAF_{II}250$, etc.), methyltransferase (ASH1, RIZ, MLL), chromatin remodelers (SNF2, Brama, CHARC, etc.) (de la Cruz et al., 2005) or other proteins (Brd2,etc.) (Kanno et al., 2004).

Histone acetylation is almost invariably associated with activation of transcription and histone deacetylation is associated with gene repression (Kurdistani and Grunstein, 2003). This is consistent with that HATs and HDACs are associated with coactivator and corepressor respectively. It has been confirmed that the bromodomain is required for chromatin association of SWI/SNF and the SAGA complexes (Hassan et al., 2002). Thus, the regulatory role of histone acetylation may function at different steps during transcription. In addition, acetylation of H4K16 modulates both higher order chromatin structure and functional interactions between a non-histone protein and the chromatin fibre (Shogren-Knaak et al., 2006).

1.6 Histone methylation

Histone methylation is catalysed by histone methyltransferases (HMT), histone-lysine N-methyltransferase and histone-arginine N-methyltransferase, which catalyse the transfer of one to three methyl groups from the cofactor S-Adenosyl methionine to lysine and arginine residues of histone proteins. All HMTs contain an SET (Su(var)3-9, Enhancer of Zeste, Trithorax) domain with exception of the non-canonical Dot1 (Dillon et al., 2005;

Martin and Zhang, 2005; van Leeuwen et al., 2002). Histone methylation is more complex than the other modifications. First, it can occur on either lysines or arginines. Secondly, histone methylation functions in both transcription activation and repression. Thirdly, lysines can be mono- (me1), di- (me2), or tri- (me3) methylated, and arginines can be mono- (me1) or di- (me2, symmetric or asymmetric) methylated. In the following paragraphs, the most recent characteristics of the individual methylation state are summarized. H3K36 methylation will be discussed in 1.8.

1.6.1 H3K4 methylation

Global patterns: The H3K4 residue in yeast is methylated by the Set1 complex across the entire ORF of active genes. As shown in Figure 6, monomethylation is enriched toward the 3' end, and dimethylation peaks in the middle, whereas trimethylation occurs around the transcription start site and the 5' end of the ORF (Pokholok et al., 2005). In vertebrates, the majority of dimethylation colocalizes with H3K4me3 in discrete zones about 5-20 nucleosomes in length proximate to highly transcribed genes (Bernstein et al., 2005; Schneider et al., 2004). A subset of dimethylation sites is devoid of H3K4me3 (Bernstein et al., 2005). Recent ChIP on chip study showed that H3K4me3 was enriched at both active and inactive genes throughout the genome. The signals observed for H3K4me3 were typically lower (about 3-fold) at the inactive genes than active genes but were substantially above background and located at the same position relative to the transcription start site (Guenther et al., 2007).

HMTs/Writers: Set1 is the sole enzyme responsible H3K4 methylation in yeast. In mammals, at least ten known or predicted H3K4 methyltransferases exist, including the MLL family (MLL1, MLL2, MLL3, MLL4, SET1A and SET1B), ASH1, SET7/9, SMYD3, and PRDM9 (Ruthenburg et al., 2007).

Demethylases/Erasers: In yeast, Jhd2 preferentially demethylates H3K4me3 (Liang et al., 2007). The first H3K4 demethylase discovered is LSD1, which demethylates H3K4me1 and H3K4me2, but not H3K4me3 (Shi et al., 2004). Subsequently, RBP2/JARID1A (Christensen et al., 2007; Klose et al., 2007b), PLU-1/ JARID1B (Yamane et al., 2007), SMCX/JARID1C (Iwase et al., 2007) and JARID1D (Lee et al., 2007a) were identified as H3K4 demethylases in the mammalian system.

Effectors/Readers: A number of effector proteins recognize methylated H3K4 through different domains. Many of the known H3K4 methyl readers reside within protein complexes associated with enzymatic activities operating on the chromatin template. Chd1 and BPTF are involved in ATP-dependent chromatin remodeling. BPTF recognizes H3K4me3 via PHD domain (Wysocka et al., 2006) and Chd1 recognizes H3K4me2 and H3K4me3 via its chromodomain (Flanagan et al., 2005; Pray-Grant et al., 2005). ING proteins are present in histone acetylation and deacetylation complexes and recognize H3K4me2 and H3K4me3 via their PHD domain (Doyon et al., 2006). JMJD2a, a histone demethylase, recognizes H3K4me3 and maybe also H3K4me2 via a tudor domain (Shi and Whetstine, 2007).

Functional significance: H3K4 methylation is coupled to specific biological functions according to its associated factors. The recruitment of remodeling machinery, such as the BPTF-containing NURF remodeling complex, facilitates transcription by increasing the accessibility of the chromatin template to the transcription machinery. The association of H3K4 methyl with ING3-5 containing acetyltransferase complexes functions in transcription activation. In contrast, ING2, a native subunit of a repressive mSin3a-HDAC1 histone deacetylase complex, binds with high affinity to the trimethylated H3K4, which functions in active gene repression. JMJD2A is present in co-repressor complex N-CoR and retinoblastoma (Huang et al., 2006). Recruitment of JMJD2A to H3K4me3 presumably leads to gene repression.

1.6.2 H3K79 methylation

Global patterns: H3K79 methylation was first identified in yeast (van Leeuwen et al., 2002). K79 lies within the core of the nucleosome rather than in the tail. Global analysis in yeast has shown that H3K79 is methylated in euchromatic regions and associates with the coding region of active genes. H3K79 trimethylation is enriched at both 5’ and 3’ across the transcribe region of genes (Pokholok et al., 2005). However, H3K79 trimethylation of the mammalian *PABPC1* gene is enriched at 5’ end and not maintained across the entire transcribed region (Vakoc et al., 2006). Global analysis on human genes indicated that H3K79 me2 is closely associated with H3K4me2 and H3K36me2 in the coding region, suggesting that active lysine methylation pairs that generally occur together in the coding regions of human genes are H3K4me2/H3K79me2 and

H3K36me2/H3K79me2. Relatively few numbers of genes appear to have H3K4me2/K36me2/K79me2 in the coding region (Miao and Natarajan, 2005).

HMTs/Writers: Dot1 was identified as the sole enzyme for H3K79 methylation in yeast (van Leeuwen et al., 2002). Conserved homologs have been subsequently identified in mammals and fly (Feng et al., 2002; Shanower et al., 2005). Dot1 is found the only active histone methyltransferase without containing SET domain so far.

Demethylases/Erasers: No demethylase for H3K79 has been identified until now.

Effectors/Readers: 53BP1 has been identified as the sole H3K79 methyl binding protein (Huyen et al., 2004).

Functional significance: The function of H3K79 methylation in transcription is contradictory. Early reports showed that H3K79me3 did not show a correlation with either active or silent genes in yeast (Pokholok et al., 2005), while H3K79me2 was linked to active transcription in *Drosophila* (Schubeler et al., 2004) and humans (Okada et al., 2005). Recent data showed that H3K79me1 was modestly associated with activation while H3K79me3 was associated with repression in human cells and that H3K79me2 did not show any preference toward either active or silent genes (Barski et al., 2007). 53BP binding to methylated H3K79me3 functions in the DNA repair process (Wang et al., 2002b).

1.6.3 H3K9 methylation

Global patterns: H3K9 methylation was first observed in mammals and the enzyme SUV39H1 was the first histone methyltransferase to be identified (Rea et al., 2000). Subsequent studies showed that H3K9 methylation, mainly H3K9me2 and H3K9me3, were enriched in heterochromatin both in yeast (Nakayama et al., 2001; Noma et al., 2001) and in mammals (Bannister et al., 2001; Lachner et al., 2001). Methylation of H3K9 has also been detected at the promoter of some mammalian genes in euchromatin region when the genes are silent. Recently, genome wide profiling has show that the signals of both H3K9me2 and H3K9me3 were higher in silent genes than active genes in a region of 10kb surrounding the TSS (transcriptional start site). High H3K9me1 levels were detected in more active promoters surrounding the TSS (Barski et al., 2007). In addition, H3K9me3 was found to be enriched also in the coding region of active genes (Vakoc et al., 2005; Vakoc et al., 2006).

HMTs/Writers: Until now, a number of H3K9 methyltransferases from different species have been identified, including Clr4 (*S.pombe*), Su(var)3-9, Ash1 (*Drosophila*) and Suv39h, G9a, Eu-HMTaseI, Setdb1 (Mammals).

Demethylases/Erasers: The JHDM3/JMJD2A family of jmjC domain containing proteins has been identified as H3K9 demethylase in mammals. This family includes JHMD3A/JMJD2A, JMJD2B, JMJD2C/GASC1, and JMJD2D (Klose et al., 2006). Rph1, the yeast homolog of mammalian JHDM3A/JMJD2A, showed demethylation activity towards H3K9me3 and H3K36me3 *in vitro* or when overexpressed in mammalian cells, although H3K9 methylation has not been found in budding yeast (Klose et al., 2007a).

Effectors/Readers: The so-called chromodomain is the first identified structure motif that allows binding to methylated histone tails. All three isoforms of HP1 (HP1α, HP1β, and HP1γ) can bind to all three methylation states of H3K9 (me1, me2, and me3) via its chromodomain. The chromodomain of CDY1 (Chromodomain protein, Y-linked, 1) binds to H3K9me2 and H3K9me3 (Kim et al., 2006).

Functional significance: As discussed above, H3K9 methylation is required for heterochromatin formation. SUV39H (or Clr4 in fission yeast) methylates H3K9, creating a binding platform for HP1 (of Swi6 in fission yeast). Upon HP1 binding, heterochromatin can spread onto adjacent nucleosomes by its association with SUV39H. In addition, HP1 self-associates via the chromoshadow domain, facilitating the spread of heterochromatin (Allis et al., 2006). H3K9 methylation also functions in the repression of euchromatic genes. The mechanism of this repression is different from that of heterochromatin formation. Generally, repressor or corepressor proteins recruit SUV39H, G9a, or SETDB1 to the promoter of target genes (Ayyanathan et al., 2003; Smallwood et al., 2007). The recent detection of H3K9me3 and HP1γ on the coding region of active genes indicates that H3K9 trimethylation may function in transcription elongation (Vakoc et al., 2005; Vakoc et al., 2006), similar to the co-occurring modification H3K36me3 on the coding region.

1.6.4 H3K27 methylation

Global patterns: H3K27 methylation is a repressive modification found in three distinct places in the cell: (1) euchromatic gene loci (H3K27me1, me2, and me3) (2) at

pericentric heterochromatin (H3K27me1), and (3) at the inactive X in mammals (H3K27me3) (Allis et al., 2006). High-resolution profiling of histone methylations in the human genome showed H3K27me3 signals were modestly elevated at silent promoter and reduced at active promoters and genic regions, whereas not much change was observed in intergenic regions. H3K27me1 signals were higher at active promoters than silent promoters, particularly downstream the TSS. H3K27me2 had a similar distribution as H3K27me3, though less biased toward silent genes (Barski et al., 2007). H3K27me1 is selectively enriched ate pericentric heterochromatin together with H3K9me3.

HMTs/Writers: The enzyme that mediates H3K27 methylation in mammals is Ezh2, a homolog of the Drosophila Enhancer of Zeste (E(z)) protein (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Muller et al., 2002). It functions in the Polycomb repressive complex 2 together with Suz12 and Eed that are required for enzymatic activity *in vivo* (Cao and Zhang, 2004; Montgomery et al., 2005).

Demethylases/Erasers: Recently, UTX and JMJD3 have been identified as H3K27me3 demethylases (Agger et al., 2007; Lee et al., 2007b).

Effectors/Readers: It has been shown that Polycomb (Pc), a subunit of PRC1 complex, binds to methylated H3K27 through its chromodomain that binds trimethyl about five times stronger than mono- or dimethylated residues (Daniel et al., 2005). Pc is encoded by a single gene in *Drosophila*. However, among five mouse homologs (Cbx2, Cbx4, Cbx6, Cbx7, Cbx8), only Cbx2 and Cbx7 bind to H3K27me3, as well as H3K9me3 (Bernstein et al., 2006b).

Functional significance: H3K27 methylation functions in (1) maintaining repression of target genes and allow for “cellular memory” throughout subsequent cell divisions and development (Ringrose and Paro, 2004), (2) heterochromatin structure formation (Peters et al., 2003), (3) X-inactivation (Wang et al., 2001), and (4) genomic imprinting (Wang et al., 2001).

1.6.5 H4K20 methylation

Global patterns: H4K20me3 is only at present pericentric heterochromatin whereas H4K20me2 is broadly distributed over euchromatic regions. H4K20me1 in enriched at Xi

foci in female cells (Schotta et al., 2004). High-resolution profiling in the human genome showed that H4K20me3 did not show association with either active or silent promoters. It also revealed a peak of H4K20me1 in the region downstream of the TSS (Barski et al., 2007). However, other data showed that H4K20me1 was enriched in promoter or coding regions of active genes (Talasz et al., 2005; Vakoc et al., 2006).

HMTs/Writers: Suv4-20h1 and Suv20h2 are responsible for generating H4H20me3 (Schotta et al., 2004). PrSet7 is an HMT with selective activity towards nucleosomal H4K20, being an exclusively monomethylating enzyme (Karachentsev et al., 2005; Nishioka et al., 2002).

Demethylases/Erasers: The demethylase of H4K20 has not been found yet.

Effectors/Readers: L3MBTL1 binds to mono- and dimethylation of histone H4 lysine 20 through its MBT domain (Trojer et al., 2007). CrB2 is H4K20 methyl binding protein in budding yeast (Sanders et al., 2004).

Functional significance: The distribution patterns and the effector proteins of H4K20 methylation indicates that this methylation may function in different processes including heterochromatin formation (Schotta et al., 2004), transcription regulation (Talasz et al., 2005; Vakoc et al., 2006), X-inactivation (Kohlmaier et al., 2004) and DNA repair (Sanders et al., 2004). H4K20 monomethylation has a negative correlation with H4K16 acetylation (Nishioka et al., 2002), which controls chromatin structure and protein interaction (Shogren-Knaak et al., 2006), suggesting that H4K20 methylation might function in higher-order chromatin structure.

1.7 DNA methylation

DNA methylation is found in the genome of diverse organisms including both prokaryotes and eukaryotes. In mammals it occurs mainly in the context of CpG dinucleotides (CpGs). Most CpGs of mammalian genome (intergenic DNA, coding DNA and repeat element) appears to be methylated, except CpGs grouped in clusters called “CpG islands” present in promoter of many genes. A subset of CpG islands has been shown to be subject to tissue-specific methylation during development.

Four DNA methyltransferases (DNMTs) sharing a conserved DNMT domain have been identified in mammals. The founding member, DNMT1, maintains DNA methylation during replication by copying the DNA methylation of the old DNA strand onto the newly synthesized strand (Leonhardt et al., 1992). DNMT3a and DNMT3b are responsible for *de novo* methylation, as they are able to target unmethylated CpG sites (Okano et al., 1999). They also cooperate with DNMT1 to propagate methylation patterns during cell division (Liang et al., 2002). DNMT2 has only weak DNA methyltransferase activity *in vitro* and has recently been shown to efficiently methylate tRNAs (Goll et al., 2006). DNMT3L is a DNMT-related protein that does not contain intrinsic DNA methyltransferase activity, but physically associates with DNMT3a and DNMT3b and modulates their catalytic activity (Hata et al., 2002). In combination, these *de novo* and maintenance methyltransferases seem to constitute the core enzymatic components of the DNA methylation system in mammals (Klose and Bird, 2006; Weber and Schubeler, 2007).

Recently, Gadd45a has been identified as a key regulator of active DNA demethylation (Barreto et al., 2007). It interacts with and requires the DNA repair endonuclease XPG, relieving epigenetic gene silencing by promoting DNA repair, which erases methylation marker.

DNA methylation is generally associated with a repressed chromatin state and inhibition of promoter activity. Two models of repression have been proposed: first, cytosine methylation can prevent binding of transcription factors; and second, DNA methylation can affect chromatin states indirectly through the recruitment of methyl-CpG-binding proteins (MBPs) (Klose and Bird, 2006). With the exception of MBD3, which contains amino acid substitutions that prevent binding to methyl-CpG, the mammalian MBD proteins (named MBD1–MBD4) and the founding member, MeCP2, all specifically recognize methyl-CpG. (Klose and Bird, 2006; Weber and Schubeler, 2007).

Mammalian DNA methylation has been implicated in a diverse range of cellular functions and pathologies, including tissue-specific gene expression, cell differentiation, genomic imprinting, X chromosome inactivation, regulation of chromatin structure, carcinogenesis, and aging. It is essential for normal development and remains indispensable for the survival of differentiated cells. The DNA methylome also

undergoes characteristic changes in pathologies such as cancer. These include genome wide loss of methylation and aberrant local gain of methylation. In particular, tumor suppressor gene promoters are targets of hypermethylation, which typically results in their silencing (Bernstein et al., 2007).

1.8 H3K36 methylation

Global patterns: H3K36 methylation was first discovered in yeast (Strahl et al., 2002). It is conserved from yeast to human, being present in *S. cerevisiae*, *S. pombe* (Morris et al., 2005), *N. crassa* (Adhvaryu et al., 2005), *C. elegans* (Bender et al., 2006), *A.thaliana* (Zhao et al., 2005), *Drosophila* (Stabell et al., 2007), mouse (Rayasam et al., 2003) and human (Sun et al., 2005). In yeast, it has been shown that mono-, di-, and trimethylation can be regulated differently although they are catalyzed by the same enzyme. In yeast, *Bur1* or *Bur2* deletion significantly reduced the level of trimethylation without apparent changes on mono- and dimethylation of H3K36 (Chu et al., 2006). Recent genome wide analyses have shown that both di- and trimethylation are enriched at the 3' open reading frame (ORF) of yeast gene (Pokholok et al., 2005). Besides of presence on 3' ORF, there is also evidence in yeast that H3K36 methylation can occur on the promoter of some genes (Morillon et al., 2005). In C. elegans, H3K36me2 has been shown to localize to autosomes, being excluded from the X chromosome (Bender et al., 2006). In fly, H3K36me2 showed an interband staining pattern on euchromatin region of polytene chromosomes (Ebert et al., 2006). In mammals, trimethylation of H3K36 is strongly enriched across the transcribed regions of active genes and H3K36me1 showed slight preference towards the active promoter (Barski et al., 2007; Vakoc et al., 2006). Acetylation of H3K36 has recently been identified as a conserved modification (Morris et al., 2007), adding complexity to the modifying state of H3K36.

HMTs/Writers: The first H3K36 methyltransferase identified is *S. cerevisiae* Set2 (Strahl et al., 2002). Set2 is the only H3K36 methyltransferase in *S. cerevisiae*, *S. pombe*, and *N. crassa* (Adhvaryu et al., 2005; Morris et al., 2005). In *C. elegans*, Mes-4 was identified as H3K36me2 methyltransferase in the germline and in the early embryo (Bender et al., 2006) and Met1 is an H3K36me3 methyltransferase (Andersen and Horvitz, 2007). In *Drosophila*, DmMes4 mediates dimethylation of H3K36, which serves as substrate for trimethylation by dHypB (Bell et al, manuscript in preparation).

However, another group reported that dSet2, an alternative name of dHypB, is the sole enzyme responsible for dimethylation of H3K36 (Stabell et al., 2007). In addition, it has been reported Drosophila discs absent, small, or homeotic-1 (ASH1) is a methyltransferase specific for H3K36 (Tanaka et al., 2007). In mammals, a number of SET domain protein have been identified as H3K36 methyltransferase, including Nsd1, Smyd2, Metnase, HypB and Ash1 (Brown et al., 2006; Lee et al., 2005; Rayasam et al., 2003; Sun et al., 2005; Tanaka et al., 2007). The structure feature of a number of SET2 family proteins is illustrated in Figure 8.

Demethylases/Eraser: Several H3K36 demethylases have been identified that contribute to the dynamics of H3K36 methylation. The first identified H3K36 demethylase is JHDM1 that selectively demethylates H3K36 with preference for the dimethyl form (Tsukada et al., 2006). Subsequently, JHDM3A (JMJD2A) was found to demethylate H3K9me3 and H3K36me3 (Whetstine et al., 2006). Recently, Jhd1 and Rph1, the yeast homolog of JHDM1 and JHDM3, were identified. Jhd1 fine-tunes the distribution of H3K36me2 within ORF (Fang et al., 2007) and Rph1 demethylates H3K36me3, H3K36me2, and also H3K9me3 (Kim and Buratowski, 2007; Klose et al., 2007a). Jhd1 and Rph1 were proposed to promote transcription elongation (Kim and Buratowski, 2007). Gis1 was reported as H3K36me1 or H3K36me2 demethylase (Tu et al., 2007). Further investigations are required to uncover the biological significance of H3K36 demethylation.

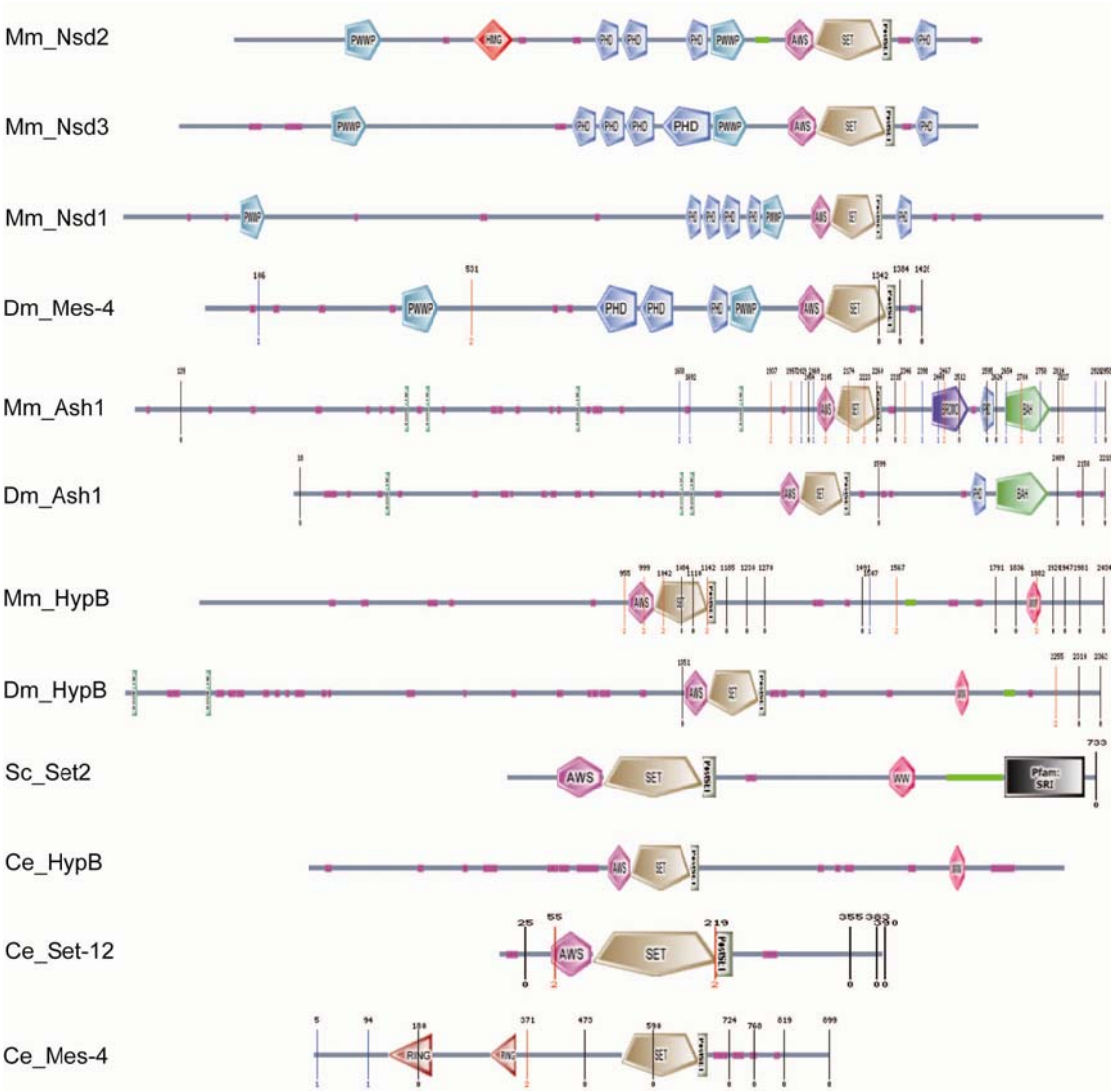

Figure 8. Structure feature of SET2 family HMTs

The SMART program was used. Different protein domains are indicated as below:

PWWP: HMG: PHD: AWS: SET:

Post-SET: Bromo: BAH: WW: Ring:

Effectors/Readers: One of the proposed models of histone modifications is that they provide binding sites that attract effector proteins (Kouzarides, 2007). It has been shown that methylation is recognized by chromo-like domains of the Royal family (Chromo, Tudor, MBT) and by PHD domains. The first effector protein binding methylated H3K36 is Eaf3, a subunit of the RPD3S complex (Joshi and Struhl, 2005; Keogh et al., 2005; Li et al., 2007b). There are two Eaf3 homologs in mammals, MSL3L and MRG15 (Carrozza et al., 2005). MSL3L is male specific lethal 3 (MSL3) like protein, a subunit of drosophila MSL3 dosage compensation complex. MRG15 is a transcription cofactor,

which is associated with both HAT and HDAC complexes. MRG15 has been shown bind to H3K36me2 or H3K36me3 *in vitro* (Zhang et al., 2006b). PHD domain containing proteins tend to recognize methylated H3K36. CHD3 (also named Mi-2α), a component of the NuRD complex, shows specific binding to H3K36me3 through its PHD domain (Mellor, 2006b). In addition, PHD domain of yeast Rco1, another subunit of RPD3S, is also required for targeting the RPD3S complex to chromatin together with the chromodomain of Eaf3. Thus, the coupled chromo and PHD domains of Rpd3S specify recognition of the methyl H3K36 mark, demonstrating the first combinatorial domain requirement within a protein complex to read a specific histone modification (Li et al., 2007b). Genome wide studies of PHD finger proteins by peptide microarray in *S. cerevisiae* have revealed two PHD finger proteins Ecm5 and Nto1 as methylated H3K36 binding proteins (Shi et al., 2007). Ecm5 is a potential histone demethylase, but no activity has been found yet (Tu et al., 2007).

Functional significance: As a conserved modification from yeast to human, H3K36 methylation plays an important role during development. In *S. cerevisiae*, synthetic growth defects were obtained when a set2 deletion was combined with deletions of all five components of the Paf1 complex in synthetic genetic array (SGA) analysis (Krogan et al., 2003). In *S. pombe*, the Set2 deletion strain showed a strong growth defect in nutrient depleted synthetic medium (Morris et al., 2005). In both cases, deletion strains show a transcription elongation defect, which is characterized by the sensitivity to 6-azauralcil, an inhibitor of transcription elongation. In *N. crassa*, H3K36 methylation is required for normal vegetable growth and sexual development. Set2 mutants grow slowly and are female sterile (Adhvaryu et al., 2005). In *C. elegans*, mutation in Mes-4 results in maternal-effect sterility (Garvin et al., 1998). Subsequent study showed that loss of Mes-4 resulted in desilencing of X-linked genes (Bender et al., 2006). In mammals, Nsd1 deficient mice are embryonic lethal (Rayasam et al., 2003). NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis (Wang et al., 2007).
All phenotypes are accompanied by misregulation of gene expression. Several lines of evidence support that H3K36 methylation functions in transcription elongation. First, in yeast SET2 deletion strains show sensitivity to 6-azauracil (Krogan et al., 2003). Second, SET2 interacts with the serine 2 phosphorylated elongating form of RNA polymerase II (Gerber and Shilatifard, 2003; Li et al., 2003; Xiao et al., 2003) and thirdly H3K36 methylation is found to accumulate at the 3' end of active genes (Pokholok et al.,

2005). Also, the methylation activity of SET2 is dependent on serine 2 phosphorylation of RNA polymerase II (Kizer et al., 2005; Krogan et al., 2003). The role of H3K36 methylation during transcription elongation, along the role of H3K4me3, is illustrated in Figure 9 (Hirose and Ohkuma, 2007).

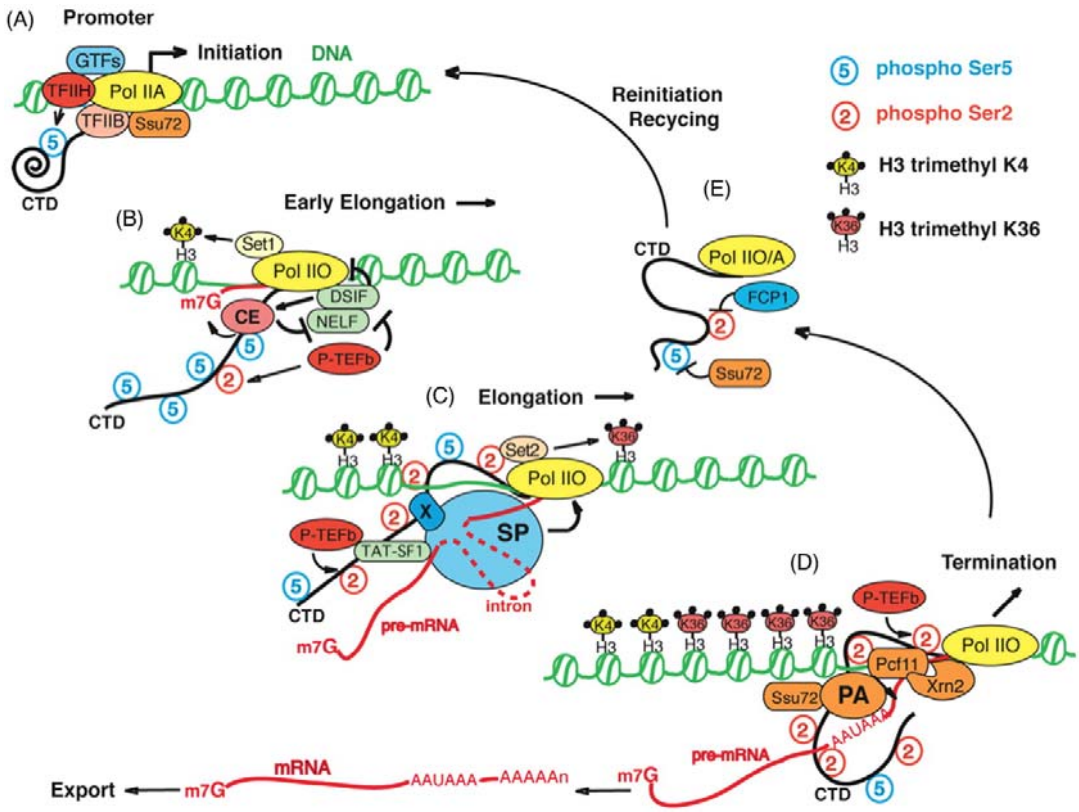

Figure 9. Function of H3K36 methylation in transcription

(A) The general transcription factors (GTFs) form a complex with initiation-competent hypo-phosphorylated Pol II (Pol IIA) at the promoter. Transcription starts at the same time as Ser5 phosphorylation of the CTD (thick black line) by TFIIH. (B) Shortly after transcription initiation, capping enzyme (CE) is recruited to the phosphorylated Pol II (Pol IIO) through its direct binding to Ser5-phosphorylated CTD. The histone methyltransferase Set1-containing complex is also recruited and trimethylates histone H3 Lysine 4 (H3K4). Transcription pausing induced by DSIF/NELF is relieved by P-TEFb-mediated CTD phosphorylation. (C) Elongating Pol IIO is increasingly phosphorylated at Ser2 by P-TEFb and associated with histone methyltransferase Set2, which trimethylates histone H3 Lysine 36. Pol IIO also helps the recruitment of the splicing machinery (SP), which splices sites in the pre-mRNA (red line). This step is mediated by an unknown phosphorylated CTD-binding factor (X) that facilitates the efficient excision of introns (red broken line). (D) Near the 3' end of the gene, 3' end processing factors (PA) are increasingly recruited to Pol IIO through direct interaction between Pcf11 and the Ser2-phosphorylated CTD. After transcribing the poly (A) signal (AATAAA), 3' end processing factors possibly transfer to RNA to catalyse endonucleolytic cleavage (black arrow) and induce subsequent transcription termination, which is presumably helped by the 5'–3' exonucleases Xrn2 and Pcf11. (E) After dissociating from the DNA template, Pol IIO is possibly dephosphorylated by the action of the CTD phosphatases, FCP1 and Ssu72, before recycling or reinitiation. (Hirose and Ohkuma, 2007)

In yeast, it was shown that the functional significance of H3K36 methylation during elongation is the suppression of inappropriate initiation from cryptic start sites within the coding region (Carrozza et al., 2005; Joshi and Struhl, 2005; Keogh et al., 2005; Li et al., 2007b; Li et al., 2007c). To achieve this, methylation of H3K36 recruits Eaf3 and Rco1 through combinatory function of chromodomain and PHD domain, which in turn brings the RPD3S deacetylase complex to the coding region. Deacetylation then removes any acetylation that was placed in the coding region during the process of transcription, resetting chromatin into its stable state. This “closing up” of chromatin, following the passage of RNA polymerase II, prevents access of internal initiation sites that may be inappropriately used (Kouzarides, 2007; Mellor, 2006a) (Figure 10). Genome-wide approach revealed that infrequently transcribed long genes exhibited a stronger dependency on Set2/Rpd3S pathway for accurate transcription (Li et al., 2007c). In mammals, the mechanistic function of H3K36 remains elusive.

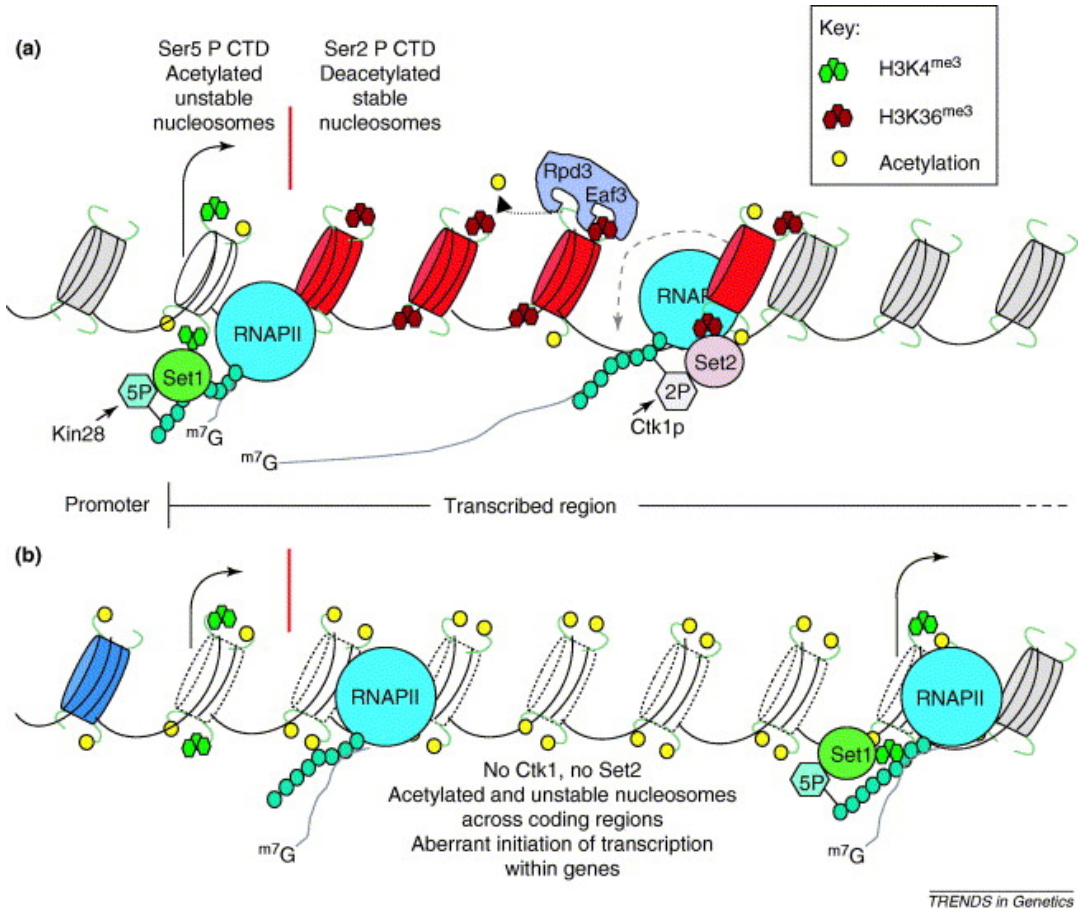

Figure 10. Mechanism of H3K36 methylation by SET2 during transcriptional elongation

The promoter and part of the transcribed region of a typical gene is shown, with associated nucleosomes, polymerases, regulatory proteins and histone modifications. The black line represents DNA and the thicker grey line RNA. Nucleosomes are discs with H3-H4 N-terminal tails coloured according to modifications (grey, unmodified; blue, acetylated at promoter; red, K36 methylated and deacetylated; white solid disc, K4 methylated and acetylated; discs with a dashed outline, acetylated and unstable). In wild-type cells **(a)** there are two clear phases of transcription elongation (shown by vertical red line in a and b). The first, after transcription initiation (bent arrow), is characterized by phosphorylation of the carboxy-terminal (CTD) of RNA polymerase II at Ser5 and associated Set1-dependent methylation of lysine 4 leading to histone acetylation and unstable nucleosomes. The second phase is characterized by CTD phosphorylation at Ser2 that recruits Set2 leading to K36me. As RNA polymerase II passes through a nucleosome, the chromatin is transiently acetylated, and then deacetylated and stabilized by the recruitment of the Rpd3 deacetylase via the Eaf3 chromodomain to methylated K36. This prevents internal initiation of transcription from cryptic sites within genes **(b)**. (Mellor, 2006)

1.9 Genetic and epigenetic regulators of pluripotency

1.9.1 ES cell self-renewal and differentiation

Mouse embryonic stem (ES) cells can be derived from the inner cell mass (ICM) of the mouse blastocyst. ES cells and the cells of the embryonic inner cell mass are pluripotent. Pluripotency refers to the potential of a cell to differentiate into any of the three germ layers: endoderm (interior stomach lining, gastrointestinal tract, the lungs), mesoderm (muscle, bone, blood, urogenital), or ectoderm (epidermal tissues and nervous system). Pluripotent stem cells can give rise to any fetal or adult cell type. However, alone they cannot develop into a fetal or adult animal because they lack the potential to contribute to extraembryonic tissue, such as the placenta. ES cell can self-renew continuously for years if they are cultured under conditions that prevent their differentiation. Pluripotent mouse ES cells are poised to differentiate into all of the somatic cells found in the embryo itself (Niwa, 2007). Studies over the past few years have revealed that both genetic and epigenetic processes play roles in the maintenance of ES cell pluripotency and the lineage commitment of ES cell.

1.9.2 Genetic regulator of ES cell self-renewal

Signaling pathways maintaining stem cell pluripotency

Extrinsic signals such as LIF, BMP, and WNT regulate pluripotent genes and maintain pluripotency and self-renewal. LIF, leukemia inhibitory factor, prevents differentiation through the canonical JAK/STAT pathway (Cartwright et al., 2005). BMP, bone morphogenetic proteins, block lineage-specific transcription factors by inducing the expression of Id genes via the Smad pathway (Ying et al., 2003). WNT/β-catenin pathway prevents ES cell differentiation through convergence on the LIF/JAK-STAT pathway at the level of STAT3 (Hao et al., 2006).

Self-renewal by preventing differentiation and by promoting proliferation

ES cell pluripotency is maintained during self-renewal by the prevention of differentiation and the promotion of proliferation. The transcription factors Oct3/4, Nanog, and Sox2 are present in the ICM and are thought to maintain pluripotency of stem cells by preventing differentiation. Oct3/4, a POU family transcription factor encoded by Pou5f1, acts as a gatekeeper to prevent ES cell differentiation. Both

upregulation and downregulation over threshold level results in cell differentiation (Niwa et al., 2000). Nanog is an NK-2 class homeobox transcription factor expressed the pluripotent cells of the ICM. Overexpression of Nanog in mouse embryonic stem cells causes them to self-renew in the absence of LIF. In the absence of Nanog, mouse embryonic stem cells differentiate into visceral/parietal endoderm (Chambers et al., 2003; Mitsui et al., 2003). Reducing the expression of Sox2, a member of the Sox (SRY-related HMG box) family, induces ES cell differentiation (Masui et al., 2007). During self-renewal, most ES cells are in the S phase of the cell cycle, with only a few in G1 (Burdon et al., 2002). Transcription factors b-Myb and c-Myc promote ES cell self-renewal by activation the progression of cell cycle directly (Cartwright et al., 2005; Iwai et al., 2001). Eras and Tcl1 promote cell cycle through modulating PI3K/Akt pathway (Takahashi et al., 2005). Utf1 (Nishimoto et al., 2005; Okuda et al., 1998) and Sall4 (Sakaki-Yumoto et al., 2006; Zhang et al., 2006a) required for cell proliferation were proved to modulate ES cell pluripotency. Recent loss- and gain-of-function screen have identified a number of candidates, such as Esrrb, Tbx3, Tcl1 and Dppa4 involved in controlling ES cell self-renewal (Ivanova et al., 2006) and the gene list keeps growing.

Transcription network for self-renewal and pluripotency

To date, ES specific enhancers that contain binding sites for Oct4 and Sox2 have been identified in several genes, including *Fgf4*, *Utf1*, *Fbxo15*, *Nanog* and *Lefty1*. Both *Oct4* and *Sox2* are activated by the Oct4-Sox2 complex in a stem cell specific manner (Chew et al., 2005). The identification of common target sites in the regulatory elements of *Oct4*, *Sox2*, and *Nanog* by recent studies using ChIP together with genome-wide location techniques has suggested that Oct4, Sox2, and Nanog might form a regulatory feedback circuit to maintain pluripotency in ES cells; in this circuit, all three transcription factors regulate themselves, as well as each other (Niwa, 2007).

To date, studies have revealed three regulatory elements of *Oct3/4*, a distal enhancer, a proximal enhancer, and proximal promoter, which can be antagonistically bound by positive or negative regulators according to undifferentiated or differentiated state of cells (Niwa, 2007) (Figure 11.A). A nuclear hormone receptor Lrh1 (Liver receptor homolog 1, also known as Nr5a2) is positive regulator of *Oct4*. By contrast, germ cell nuclear factor (Gcnf, or Nr6a1) is a negative regulator of *Oct4*. Chicken ovalbumin upstream promoter-transcription factors (Coup-tf) I and II, encoded by *Nr2f1* and *Nr2f2*,

respectively, also function as negative regulators of *Oct3/4* expression. The balance between these positive and negative regulators might determine the precise level of Oct3/4 expression in response to extracellular stimuli.

The regulatory circuit that maintains pluripotency interacts with the feedback loop shown in figure 11.B, in which Oct3/4, Sox2 and Nanog function to maintain their expression, promoting continuous ES cell self-renewal. This loop determines the differentiation fate of ES cells by influencing the expression of transcription factors, such as Cdx2 (which promotes trophectodermal differentiation) and Gata6 (which promotes primitive endoderm differentiation). Moreover, as Gcnf, Nr2f1 and Nr2f2 are upregulated after the induction of either trophectoderm or primitive endoderm differentiation, these negative regulators might form the negative-feedback loop that shuts down *Oct3/4* in differentiated cells, and which could then be followed by epigenetic chromatin modifications that result in the repression of the *Oct3/4* promoter (Niwa, 2007).

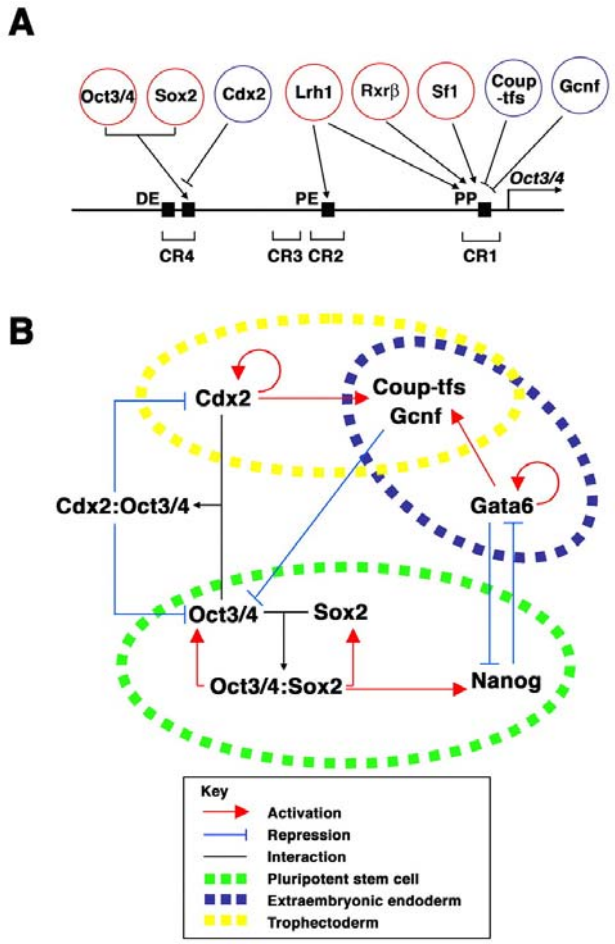

Figure 11. A transcription factor network to control ES cell self-renewal and differentiation.

(**A**) Transcriptional regulation of the mouse *Oct3/4* gene. There are four evolutionally conserved regions (CR1-4) that contain multiple transcription factor (TF) binding sites. The TFs that bind to these sites are shown above and either activate (red) or repress (blue) transcription. DE, distal enhancer; PE, proximal enhancer; PP, proximal promoter. (**B**) Transcription factor networks for pluripotent stem cells (green), trophectoderm (yellow) and primitive (extraembryonic) endoderm (blue). Positive-feedback loops between Oct3/4, Sox2 and Nanog maintain their expression to promote continuous ES cell self-renewal. Cdx2 is autoregulated and forms a reciprocal inhibitory loop with Oct3/4, which acts to establish their mutually exclusive expression patterns. A similar regulatory loop, not yet confirmed, might exist for Nanog and Gata6. Coup-tfs and Gcnf act as a negative-feedback system to repress Oct3/4 completely. (Niwa, 2007)

1.9.3 Chromatin regulators of pluripotency

A series of recent studies have revealed that mouse and human ES cells possess certain novel epigenetic features, such as chromatin structure, abundance of modified histones, and Polycomb group (PcG) protein-binding patterns (Meshorer and Misteli, 2006).

Properties of stem cell chromatin

Many nuclear features, including nuclear lamina, the nucleolus, heterochromatin structure, and nuclear speckles, undergo morphological changes during the differentiation process (Meshorer and Misteli, 2006). Studies in several systems indicate that ES cells are characterized by a distinct higher-order global chromatin structure. ES cells are richer in less compact euchromatin and, as differentiation progresses, accumulate highly condensed, transcriptionally inactive heterochromatin regions (Arney and Fisher, 2004; Francastel et al., 2000). Heterochromatin spatially rearranges and the number of heterochromatic foci increases during differentiation (Meshorer and Misteli, 2006).

Consistent with this notion, the expression of several ATP-dependent chromatin remodelling factors is elevated in ES cells and the genomic disruption of chromatin remodelling proteins, including BRG1 (Bultman et al., 2000), SNF5 (Klochendler-Yeivin et al., 2000), SSRP1 (Cao et al., 2003) and SNF2H (Stopka and Skoultchi, 2003), results in premature embryonic death prior to implantation. Interestingly, in all these cases, lethality occurred at the blastocyst stage, when the inner cell mass (ICM), the source of ES cells, is formed. It has been shown that the chromatin remodeller NuRD (nucleosome remodelling and histone deacetylation) is essential for ES cell differentiation (Kaji et al., 2006). Although the detailed mechanism of their functions is unknown, these observations strongly point towards an active role of chromatin-remodelling factors in the maintenance of stem-cell identity and the initial steps of differentiation.

Recent FRAP (fluorescent recovery after photobleaching) studies have revealed that histones H2B, H3, and HP1 are more rapidly exchanged in ES cells compared with differentiated cells (Meshorer et al., 2006). The chromatin state might be more permissive for transcription in ES cells than in differentiated cells.

Histone modification has been correlated with the chromatin and the transcriptional status of genes. Consistent with changes in the global genome activity, changes in histone-modification patterns accompany ES-cell differentiation. Examples are the differentiation-dependent increase in the silenced chromatin mark H3K9 trimethylation and a decrease in the global levels of active markers acetylated histones H3 and H4 (Lee et al., 2004; Meshorer et al., 2006) . These observations indicate that ES-cell chromatin is overall more active, or at least marked with activity-associated histone modifications, and that differentiation is accompanied by a transition to transcriptionally less-permissive chromatin. An elevation of repressive heterochromatin marks, including H3-triMeK9, H3-MeK27, H3-diMeK27, H4-diMeK20 and H4-triMeK20, was also observed in repeat sequences and retrotransposons during RA-induced mouse ES-cell differentiation (Martens et al., 2005). Global histone deacetylation during ES cell differentiation is also implied by the observed inhibition of differentiation of murine ES cells after treatment with the histone deacetylase (HDAC) inhibitor trichostatin A (TSA) (Lee et al., 2004).

Bivalent chromatin in stem cell

The properties of ES cells described above indicate that the pluripotency of ES cells is underpinned by an unusual state of their chromatin and the distinct state of chromatin is complemented by unique epigenetic mechanism to sustain pluripotency. A series of reports have assigned to Polycomb-group (PcG) proteins an essential role in maintaining the pluripotent state of ES cells as PcG complexes bind to many differentiation specific genes (Boyer et al., 2006; Lee et al., 2006). Meanwhile, genome wide mapping of H3K27me3 (a repressive marker catalysed by PcG proteins) and H3K4me3 (an active marker catalysed by Trithorax-group (TrxG) proteins) by CHIP-chip revealed an unusual “bivalent” chromatin structure, in which these two mutually exclusive markers colocalized in particular region named “bivalent domain” (Bernstein et al., 2006a; Jorgensen et al., 2006). The bivalent domains are highly enriched in ES cells relative to differentiated cells and they are associated with genes encoding transcription factors with roles in embryonic development and lineage specification. Subsequently, genome wide mapping of chromatin state by an alternative method CHIP-seq (mapping CHIP enrichment by sequencing) characterized the changes of the bivalent domain

during differentiation from ES cells to neural progenitor cells and embryonic fibroblast. The change of bivalent promoters on a set of genes, involved in different differentiation pathways, correlated to their demonstrated developmental potential (Mikkelsen et al., 2007).

Bivalent promoters show low activity despite the presence of H3K4me3, suggesting that the repressive effect of PcG activity is generally dominant over the ubiquitous TrxG activity. Bivalent domains are associated with genes with more complex expression patterns, including key developmental transcription factors, morphogens and cell surface molecules. Thus, these results suggest that promoters may be classified as active, repressed or poised for alternative developmental fates. Conceivably, chromatin state at key regulatory genes may suffice to describe developmental commitment and potential (Mikkelsen et al., 2007).

It is not understood how PcG proteins are recruited to specific target for establishing the bivalent domains in ES cells. It is known that half of the bivalent domains are also bound by Oct4, Sox2, and Nanog (Bernstein et al., 2006a). These stem-cell-specific transcription factors might take the role of recruiting PcGs to bivalent domains. Other novel stem-cell-specific transcription factors must be present in pluripotent cells and contribute to setting up chromatin signatures of stem cells. Thus, the orchestration of network architecture of ES cells and the balance between genetic and epigenetic regulation are key to understand the phenomenon of pluripotency.

Epigenetic regulation of Oct4 expression

Both global and local chromatin states are investigated to define how transcription factors and chromatin modifiers function together over time in establishing and maintaining the pluripotent state. As well as global changes, local chromatin modifications are thought to be important for the proper control of stem cell or lineage specific genes. As expected, epigenetic changes on the locus of *Oct4* during differentiation have been confirmed. It has been shown that several genes, such as *G9a*, *Mbd3*, *Suz12*, and *CGBP*, are required for repression and silencing of *Oct4* during differentiation (Carlone et al., 2005; Kaji et al., 2007; Pasini et al., 2007). In a recent study, the decrease of histone deacetylation and the increase of DNA methylation were observed on the *Oct4* promoter in ES cells induced for differentiation by addition of all-

trans retinoic acid (RA). A decrease of H3K4 methylation and increase of H3K9 methylation was also observed. *Oct4* in differentiated ES cells can be reactivated in the absence of DNA methylation (Feldman et al., 2006). One model was proposed based on this evidence (Figure 13). Nucleosomes over the active *Oct4* promoter initially contain histone H3 that is acetylated at Lys 9 and Lys 14. When differentiation is initiated, repressors (including orphan nuclear receptor GCNF, which binds the RA receptor element, RARE) associate with the *Oct4* promoter, causing transient transcriptional repression. This presumably brings about the binding of G9a, which recruits histone deacetylase molecules (HDACs) by an, as yet, uncharacterized mechanism. Once deacetylated, H3K9 becomes a substrate for methylation, either by G9a itself, or through the involvement of additional histone methylases. H3K9me3 can then bind the chromodomain protein HP1. DNA methylation is catalysed by Dnmt3a, and perhaps Dnmt3b, which are recruited to the promoter through the involvement of G9a and other effectors, such as HP1. Prior to *de novo* methylation, *Oct4* can be reactivated when cells are returned to early pre-differentiation conditions. Following this step, however, repression seems to be irreversible.

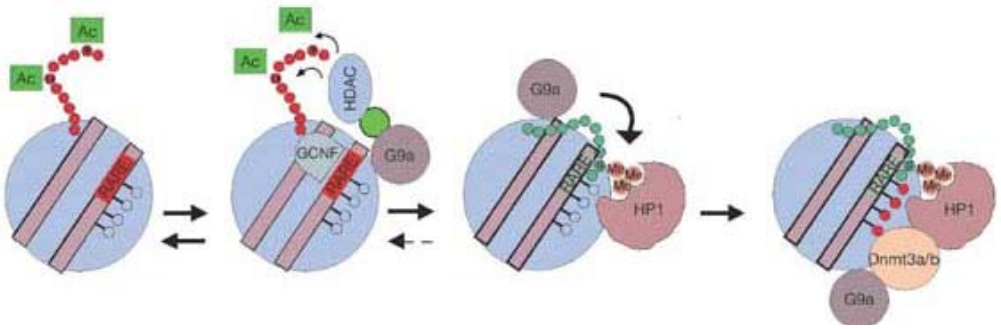

Figure 13. Model for *Oct4* heterochromatinization

(Feldman et al., 2006)

Recently, Mbd3/NuRD was found to be required for repression of *Oct4* during differentiation in embryoid bodies (Kaji et al., 2006). Other data show that GCNF sequentially recruits MBD3 and MBD2, and initiate DNA methylation on the promoter of *Oct4* in RA-differentiated ES cells, resulting in repression and silencing of *Oct4*

expression (Gu et al., 2006) (Figure 15). It is not clear if the above two model function independently or synergistically.

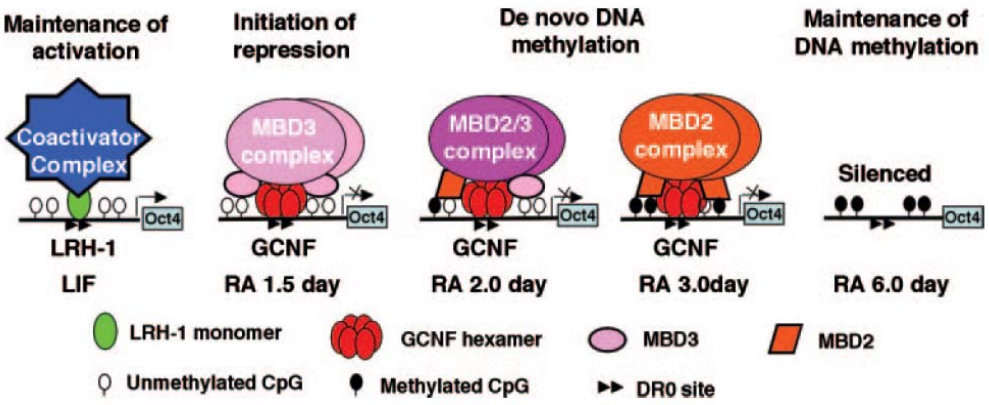

Figure 14. Model of *Oct4* gene repression and silencing initiated by GCNF-dependent recruitment of MBD2 and MBD3

The *Oct4* promoter with a doublet of direct repeats with 0 spacing (DR0) and unmethylated CpG sites is activated by LRH-1 under the control of LIF. At the beginning of RA induction (1.5 days), induced expression of GCNF hexamer replaces LRH-1 binding at the DR0 site. GCNF recruits the MBD3 complex to unmethylated CpG sites, and *Oct4* repression is initiated. Once *de novo* DNA methylation is triggered through direct or indirect recruitment of Dnmt3A or -3B to the *Oct4* promoter by the GCNF-MBD3 complex, MBD2 or MBD2/3 complexes are recruited to methylated CpG sites, and silencing of the *Oct4* gene occurs (days 1.5 to 3.0). At late stages of RA-induced differentiation (days 3 to 6), the expression of GCNF is down-regulated, and the MBD2 and MBD3 complexes are no longer bound to the *Oct4* promoter, but DNA methylation is maintained, and the *Oct4* gene is completely silenced (day 6). (Gu et al., 2006)

In this thesis, I addressed the function of H3K36 methylation in mammals. The distribution patterns of H3K36me1, me2, and me3 was analyzed in human and mouse cells by IF. The enzymatic activity of candidate H3K36 methyltransferase, Nsd and HypB proteins, was measured by HMT assay *in vitro* and RNAi approach *in vivo*. The dynamic and function of H3K36me3 during ES cell differentiation was investigated.

CHAPTER 2 RESULTS

2.1 Distribution of H3K36 methylation in mammalian cells

To investigate a possible role of H3K36 methylation in euchromatin or heterochromatin formation, the localization of H3K36 methylation states was determined by immunofluorescence staining using antibodies specific for H3K36me1, H3K36me2, and H3K36me3. The fluorochrome DAPI was used as a nuclear counterstain. In somatic cells, pericentric heterochromatin of several chromosomes clusters into chromocenters that can be visualized by DAPI preferentially binding to the underlying AT-rich major satellites (Guenatri et al., 2004). In contrast, euchromatin is only weakly stained by DAPI. The localization pattern of H3K36 methylation was analyzed in differentiated NIH3T3 cells, mouse embryo stem cells and human cells.

2.1.1 Distribution of H3K36 methylation in differentiated mammalian cells

The localization pattern of H3K36 methylation in mouse NIH3T3 cells is shown in Figure 15. Euchromatin and heterochromatin were marked by H3K36me1 and H3K36me2, as indicated by the labeling of regions stained weakly or intensively with DAPI. In contrast, H3K36 trimethylation only localized to euchromatin. H3K36 methylation states were also analyzed in human cells. All three methyl markers are evenly distributed throughout the nucleus of 293 cells (Figure 16) and U2OS cells (Figure 17).

To see if H3K36 methylation patterns are dynamic throughout the cell cycle, the localization was analyzed in G1, S, and G2/M phase NIH3T3 cells. NIH3T3 cells were fractionated by FACS sorting based on Hoechst 33342 staining (Figure 18). Proliferating cell nuclear antigen (PCNA) antibody was used as a marker to distinguish early, mid, and late S phase cells. PCNA shows a diffused pattern in non-S phase cells and a granular pattern in early S cells. In mid S cells, PCNA shows ring shape-like staining around pericentric heterochromatin or spotted foci colocalized with heterochromatin, indicating the replication of major satellites. In late S cells, PCNA displays spotted foci on the edge of pericentric heterochromatin indicating the replication of minor satellites (Guenatri et al., 2004; Houlard et al., 2006). The H3 serine10 phosphorylation antibody was used as marker for cells in G2 and mitosis. Late G2 cells show H3 serine10 phosphorylation at pericentric heterochromatin only. As mitosis proceeds,

phosphorylation of histone H3 spreads along the chromosomes and is complete at prophase. At the end of mitosis, histone H3 is dephosphorylated (Polioudaki et al., 2004; Prigent and Dimitrov, 2003). Mitotic chromosomes were identified by the morphology of its chromosomes. No apparent difference of distribution patterns for all three methyl markers, H3K36me1 (Figure 19), H3K36me2 (Figure 20), and H3K36me3 (Figure 21) was found between cells from different phase of cell cycle (compare G1, S, and G2, early S and later S, early G2 and later G2). These results indicate that genomic localization patterns of H3K36 methylation remain constant throughout the cell cycle.

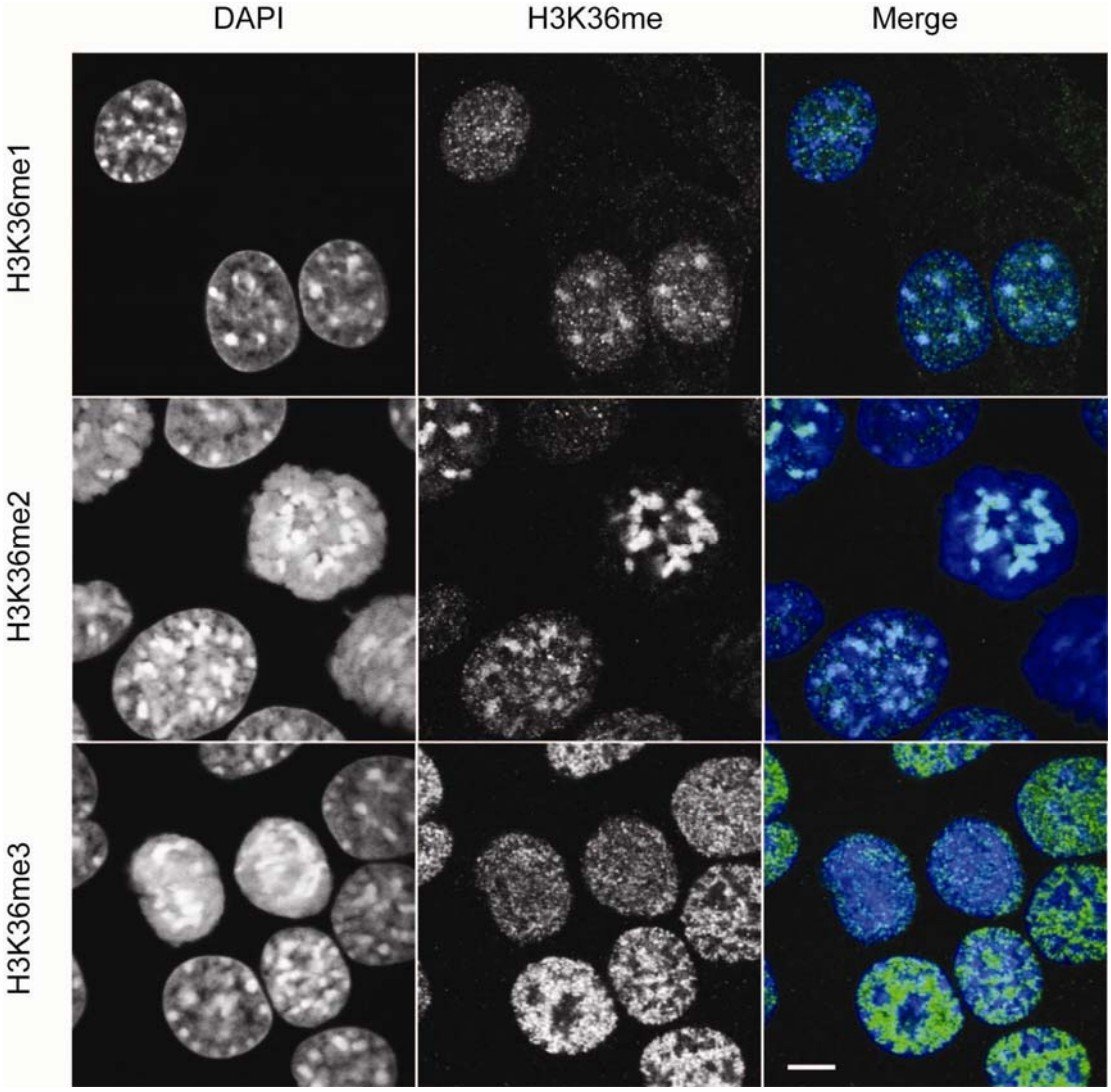

Figure 15. Distribution pattern of H3K36 methylation in mouse NIH3T3 cells

NIH3T3 cells were stained with H3K36 methylation antibodies and counterstained with DAPI. H3K36me1 and H3K36me2 localized to euchromatin and heterochromatin. H3K36me3 only localized to euchromatic regions. Scale bar: 5 micrometer.

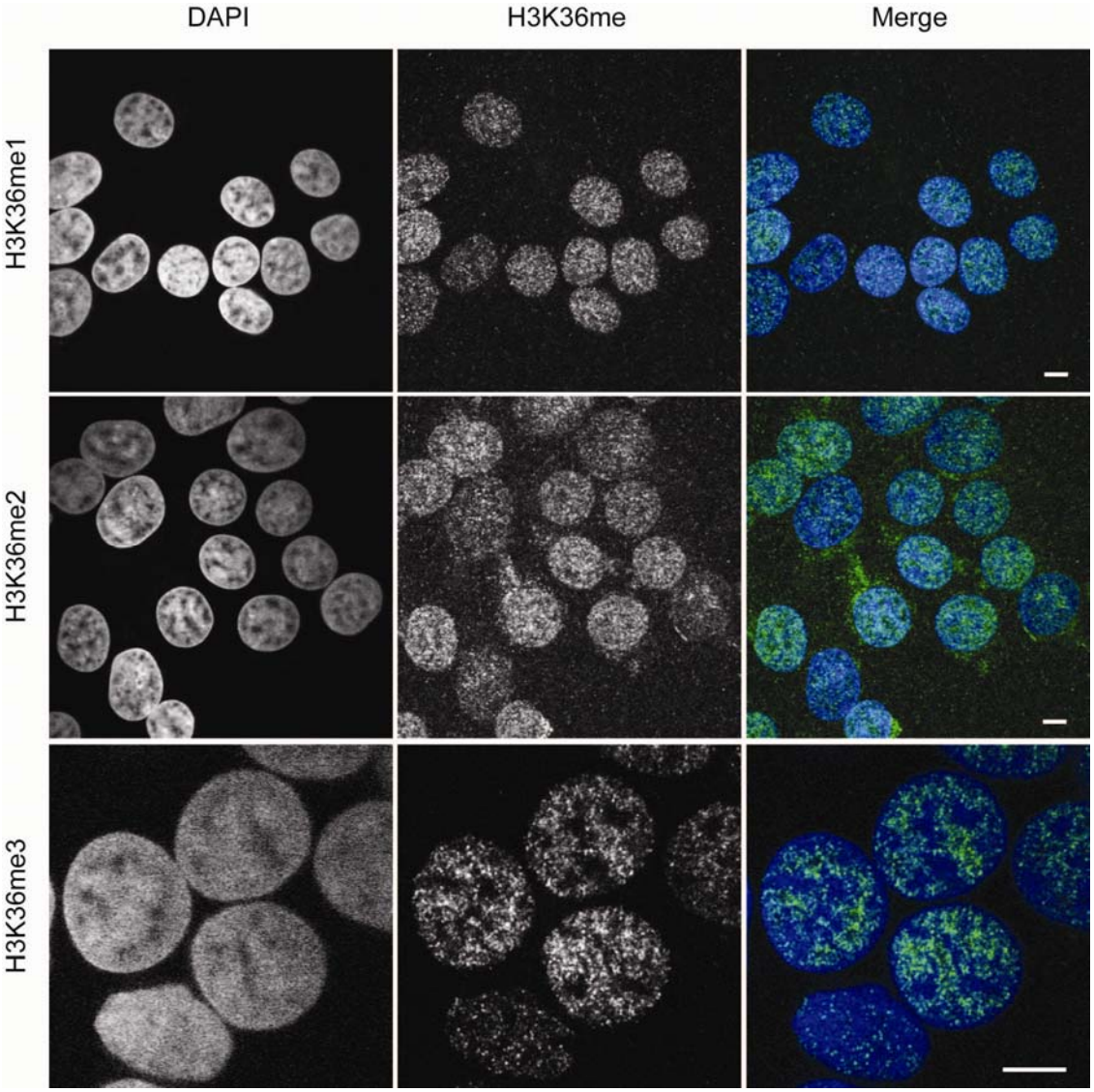

Figure 16. Distribution pattern of H3K36 methylation in human 293 cells

H3K36me1, H3K36me2, and H3K36me3 are evenly distributed throughout the nucleus in 293 cells. Human cells do not contain chromocenters, intensively stained by DAPI, that are characteristic of mouse cells. Scale bar: 5 micrometer.

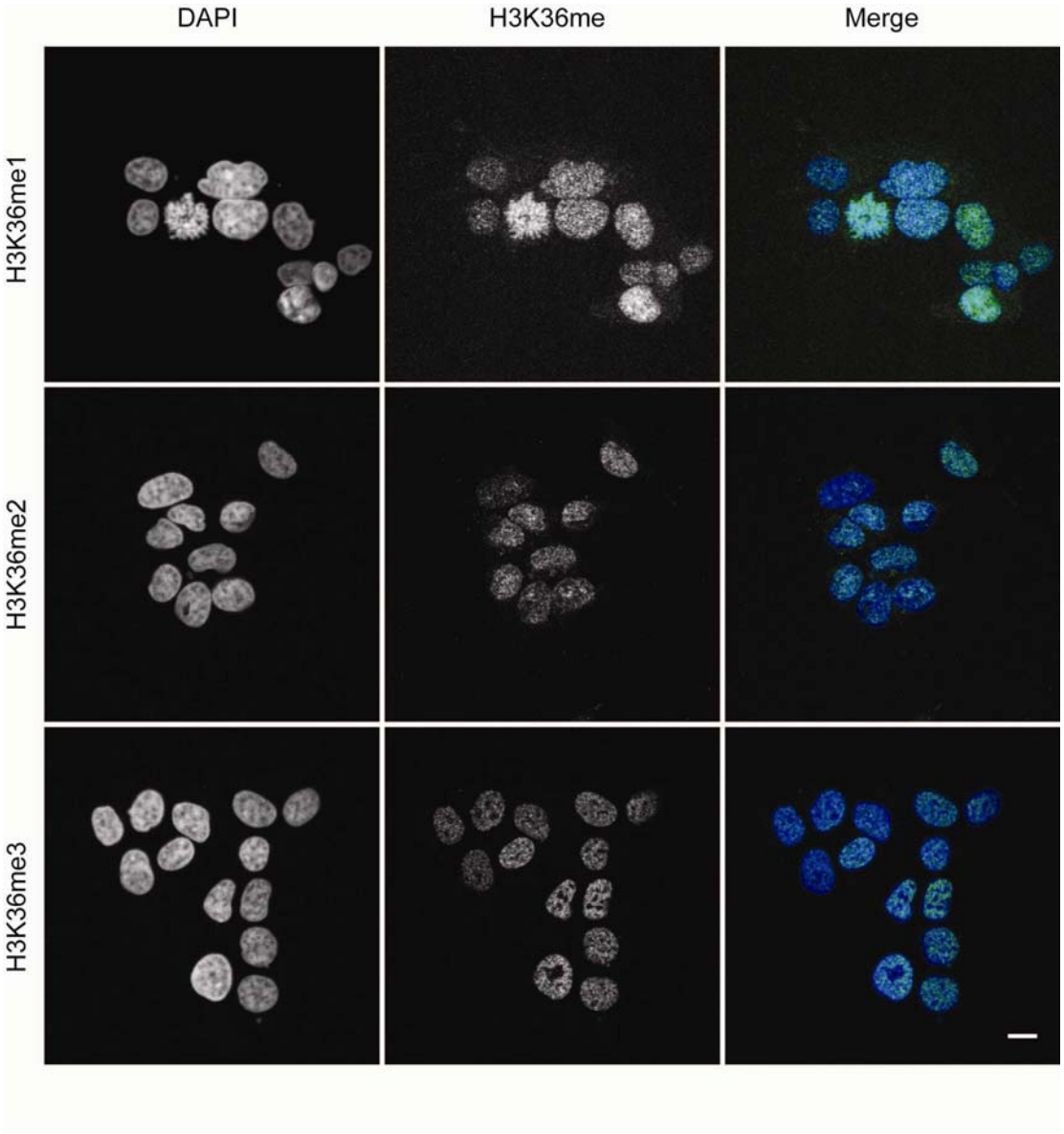

Figure 17. Distribution pattern of H3K36 methylation in U2OS cells

H3K36me1, H3K36me2, and H3K36me3 are evenly distributed throughout the nucleus in U2OS cells. Scale bar: 5 micrometer.

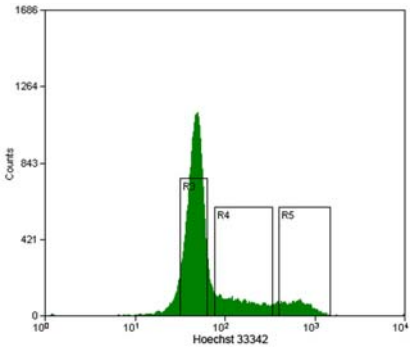

Figure 18. FACS sorting profile of NIH3T3 cells

NIH3T3 cells were sorted based on Hoechst 33342 staining. G1, S, and G2/M cells were separately collected for immunofluorescence staining.

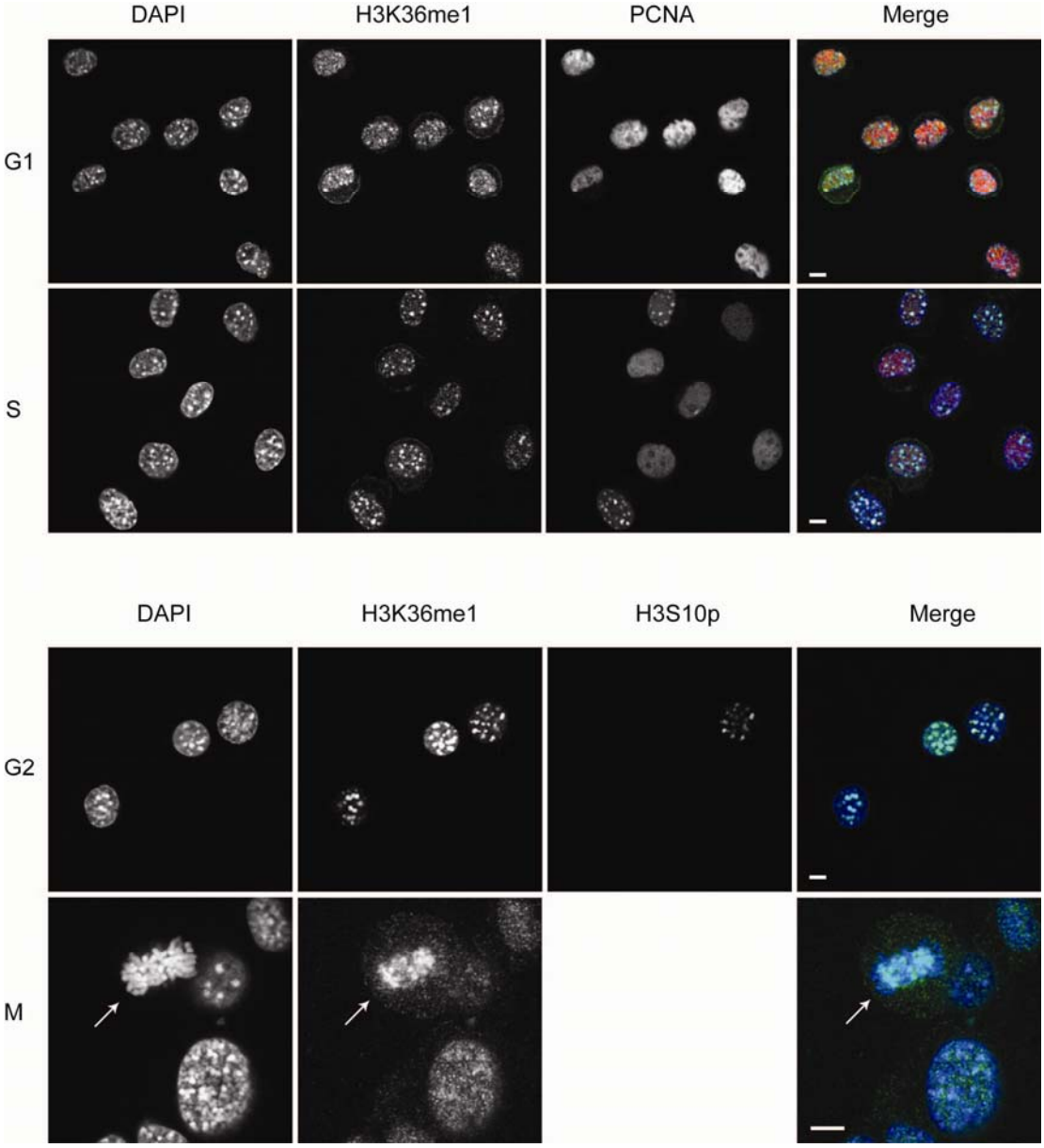

Figure 19. Distribution pattern of H3K36me1 during cell cycle in NIH3T3 cells

H3K36me1 localized to euchromatin and heterochromatin during the entire cell cycle G1, S, and G2/M phase; early S and later S are distinguished by PCNA staining; early G2 and later G2 are distinguished by H3 serine 10 phosphorylation staining. Mitotic chromosomes are identified by the morphology of chromosomes stained by DAPI. Scale bar: 5 micrometer.

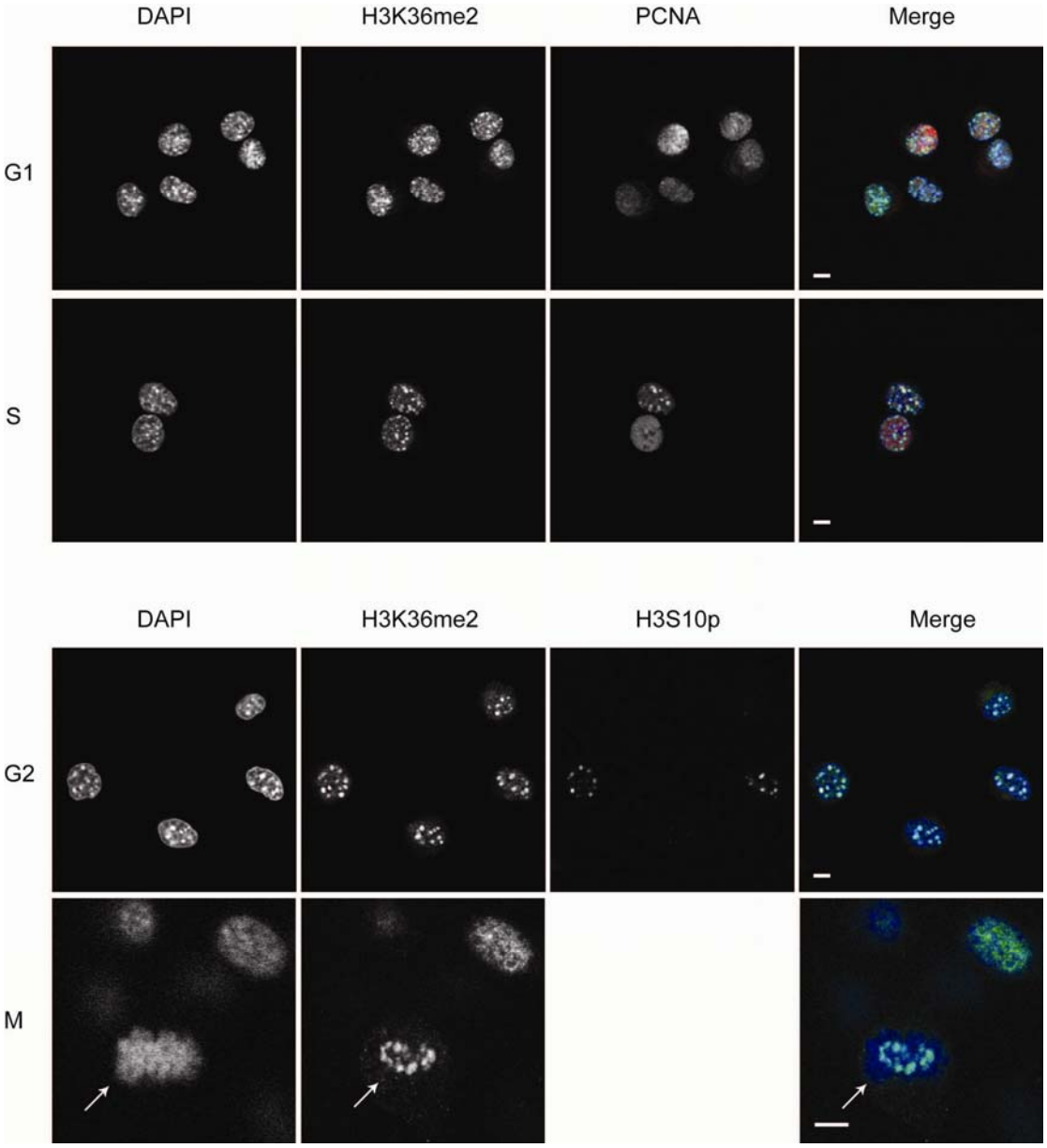

Figure 20. Distribution pattern of H3K36me2 during cell cycle in NIH3T3 cells

H3K36me2 localized to euchromatin and heterochromatin during the entire cell cycle G1, S, and G2/M phase; early S and later S are distinguished by PCNA staining; early G2 and later G2 are distinguished by H3 serine 10 phosphorylation staining. Mitotic chromosomes are identified by the morphology of chromosomes stained by DAPI. Scale bar: 5 micrometer.

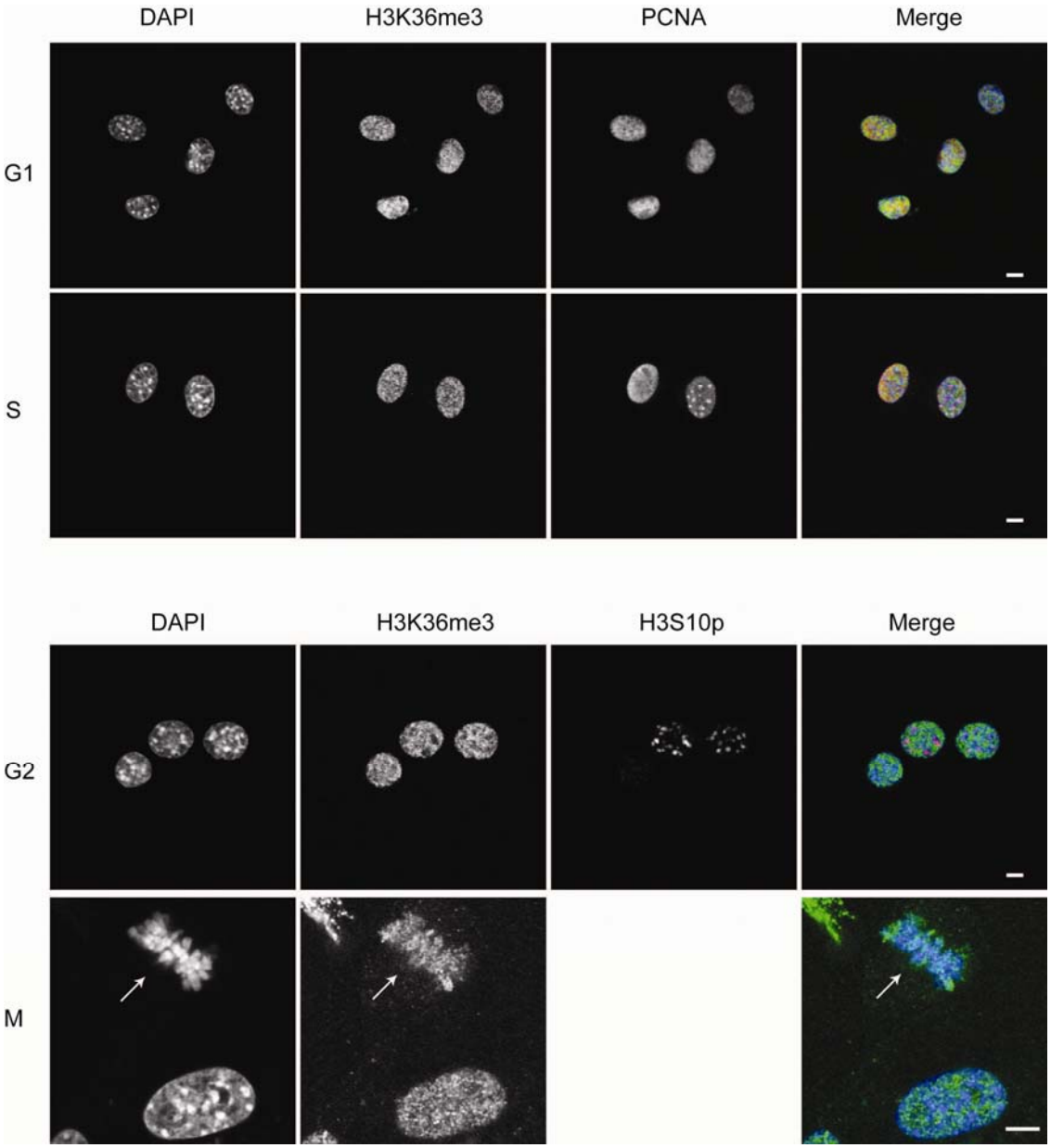

Figure 21. Distribution pattern of H3K36me3 during cell cycle in NIH3T3 cells

H3K36me3 localized to euchromatin during the entire cell cycle G1, S, and G2/M phase; early S and later S are distinguished by PCNA staining; early G2 and later G2 are distinguished by H3 serine 10 phosphorylation staining. Mitotic chromosomes are identified by the morphology of chromosomes stained by DAPI. Scale bar: 5 micrometer.

2.1.2 Distribution of H3K36 methylation in undifferentiated embryonic stem cells

It was shown that the distribution pattern of some histone methyl markers differ between undifferentiated and terminally differentiated cells (Biron et al., 2004; Kourmouli et al., 2004). To address if H3K36 methylation states change during differentiation, immunofluorescence staining was performed in CCE cells, a mouse embryonic stem cell line. An antibody against Oct4 was used to validate the undifferentiated state of the pluripotent stem cells. Intriguingly, the staining revealed a clear inverse correlation between H3K36 methylation and Oct4 expression (Figure 22). Many CCE cells, in which Oct4 was expressed at a high level, displayed only low levels of staining intensity for all three methyl states of H3K36. Only a small population of cells, in which Oct4 was lowly expressed, showed high level of intensity for all three H3K36 methylation states (Figure 22).

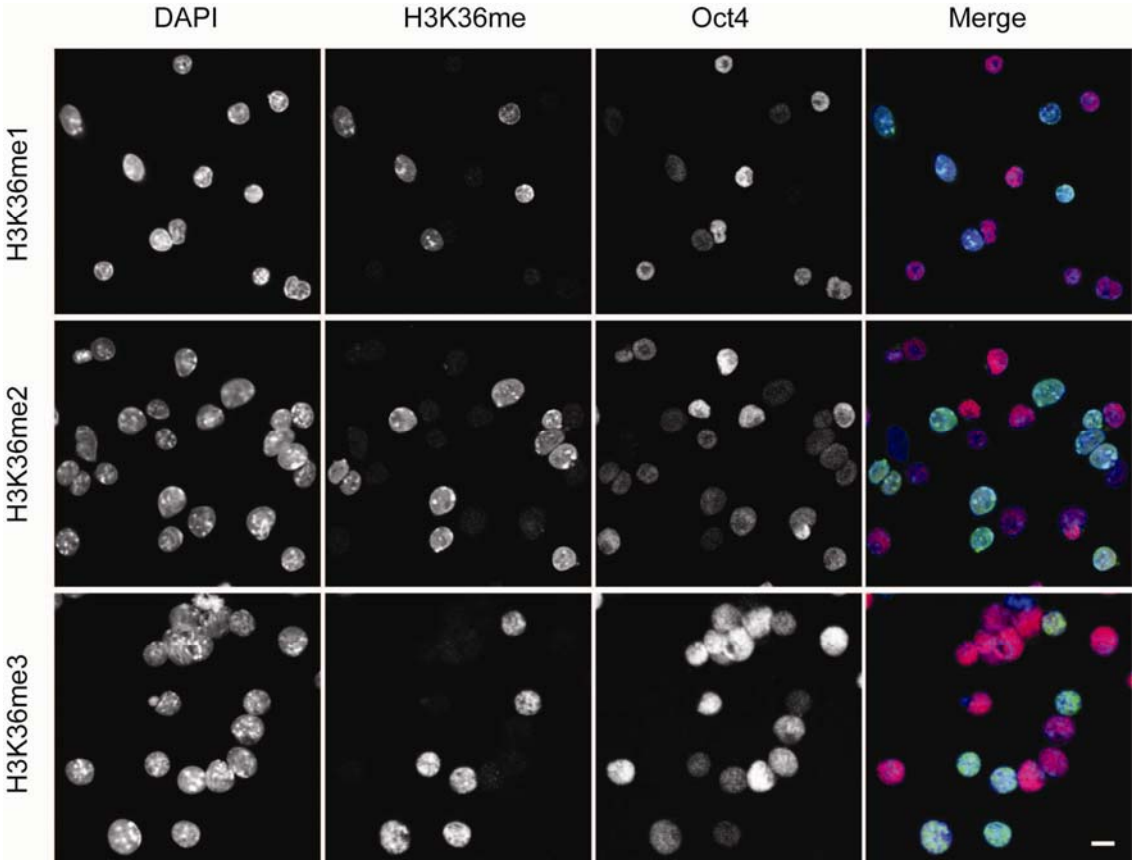

Figure 22. Inverse correlation between H3K36 methylation and Oct4 expression

Many CCE cells showed high level intensity of Oct4 but low level of intensity for all three methyl markers of H3K36. Only small population of cells, in which Oct4 was expressed at low level, showed high level of intensity for all three H3K36 methylation states, suggesting an inverse correlation between H3K36 methylation and Oct4 expression. Scale bar: 10 micrometer.

In cells with low Oct4 levels, it was noticed that H3K36me1 and H3K36me2 also showed heterochromatic foci at DAPI dense region (Figure 23). This staining pattern resembles that of NIH3T3 cells, suggesting that the CCE cells with low Oct4 levels represent a differentiated state instead of technical failure to detect Oct4. Like NIH3T3 cells, H3K36me3 only localized to euchromatin in cells with low Oct4 level. It has been reported that 5-10% CCE cells undergo spontaneous differentiation even with the presence of LIF (Sauter et al., 2005). Taken together, these data suggest an inverse correlation between H3K36 methylation and the expression of Oct4, which may point to a function of H3K36 methylation during ES cell differentiation.

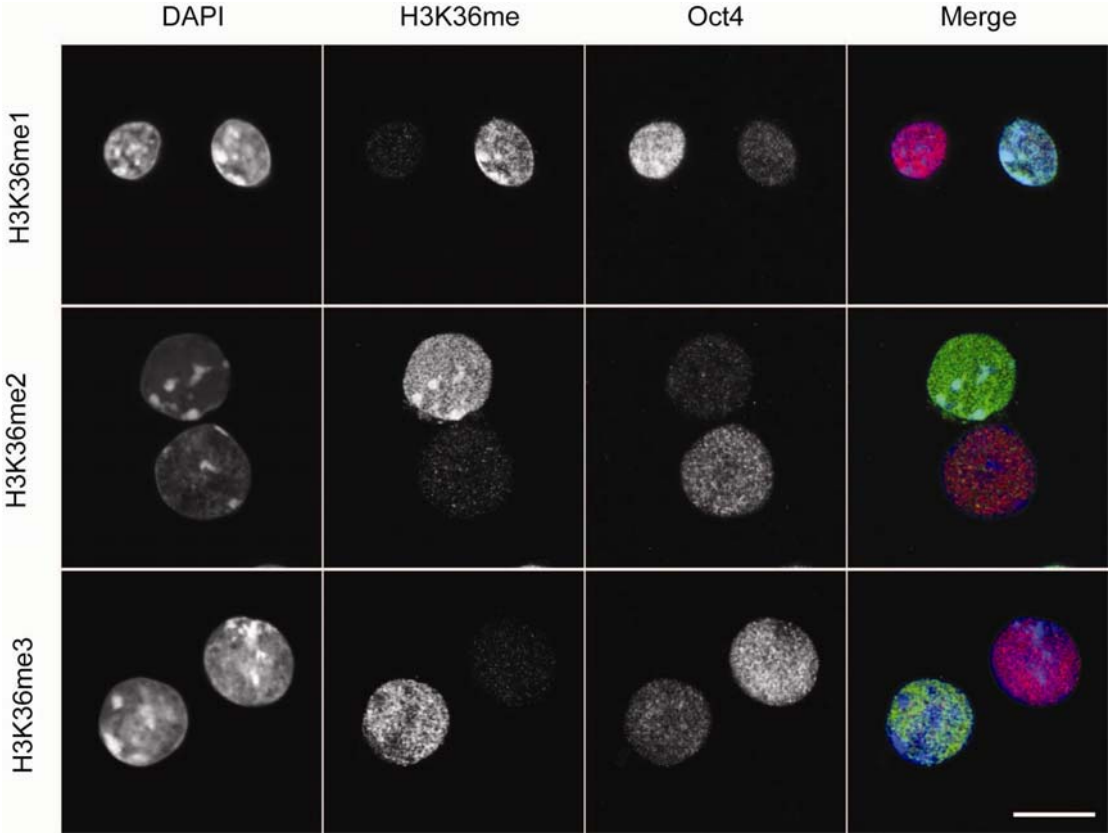

Figure 23. H3K36me1 and H3K36me2 showed heterochromatic foci in cells with low level of Oct4

H3K36me1/2 showed enrichment at heterochromatin region in cells with low level of Oct4. H3K36me3 only showed euchromatic distribution. Scale bar: 10 micrometer.

To exclude that the inverse correlation is cell cycle dependent, costaining of H3K36me3 and Oct4 was performed in G1, S, and G2/M phase of CCE cells. CCE cells were fractionated by FACS sorting to G1, S, and G2/M based on Hoechst 33342 staining. This inverse correlation H3K36me3 and Oct4 expression existed throughout the cell

cycle. The proportion of cells with high level of K36me3 and low level of Oct4 was about 10% in G1, S, and G2 phase (Figure 24). Cells in mitotic stage did not show nuclear Oct4 signals (last panel of Figure 24).

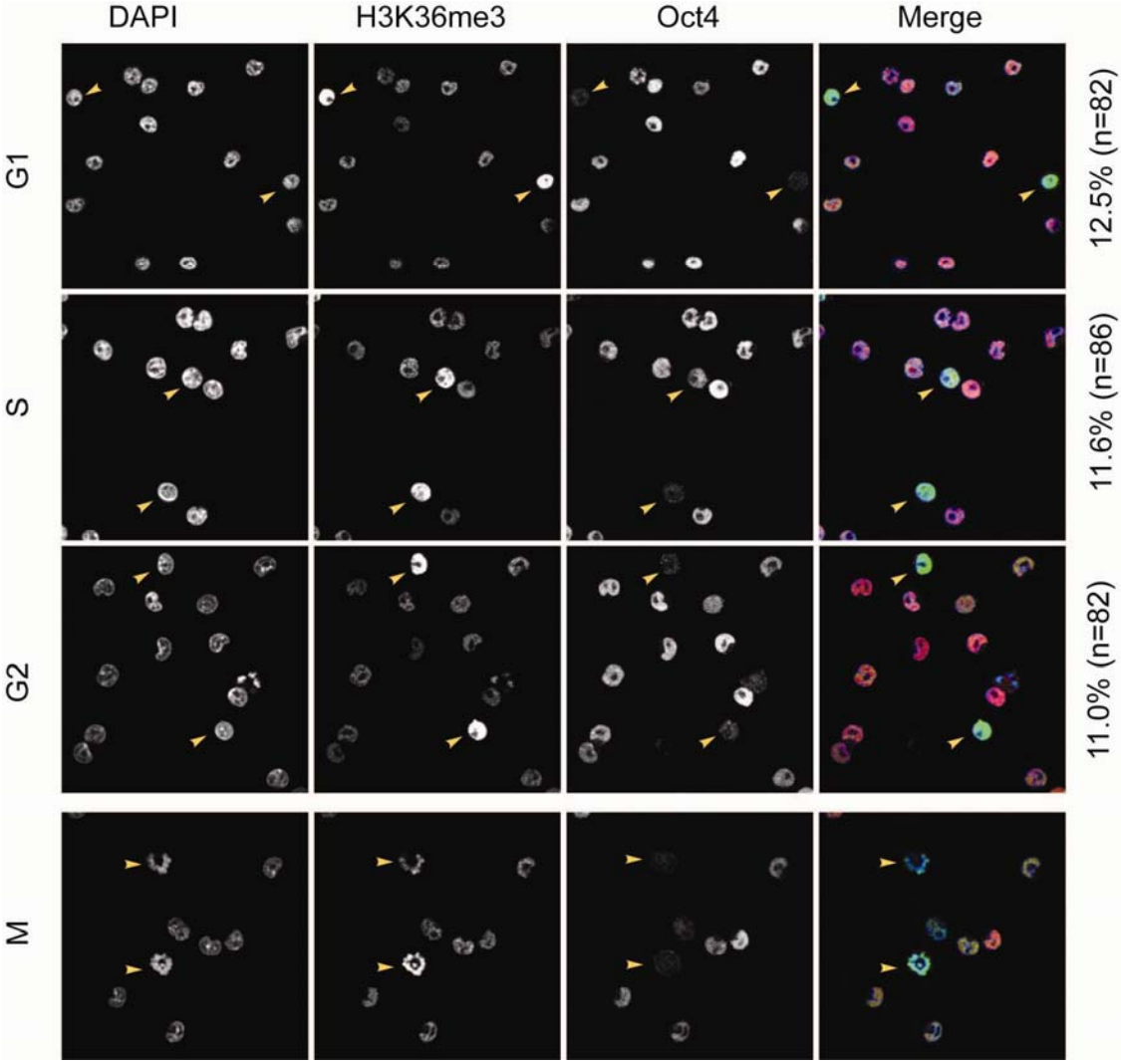

Figure 24. Inverse correlation between H3K36me3 and Oct4 expression throughout cell cycle

G1, S, and G2/M CCE cells were fractionated based on Hoechst 33342 staining and stained by H3K36me3 and Oct4 antibodies. In all three phases, about 10% of cells showed low levels of Oct4 but high levels of H3K36me3 (indicated by arrowhead). Cells in mitotic stage showed no signal of Oct4 (indicated by arrowhead of the bottom panel).

To test if this reverse correlation is universal in ES cells, immunostaining of H3K36me3 and Oct4 expression was performed in other two ES cell lines, PGK12 and LF2. These lines showed a similar inverse correlation as observed in CCE cells (Figure 25).

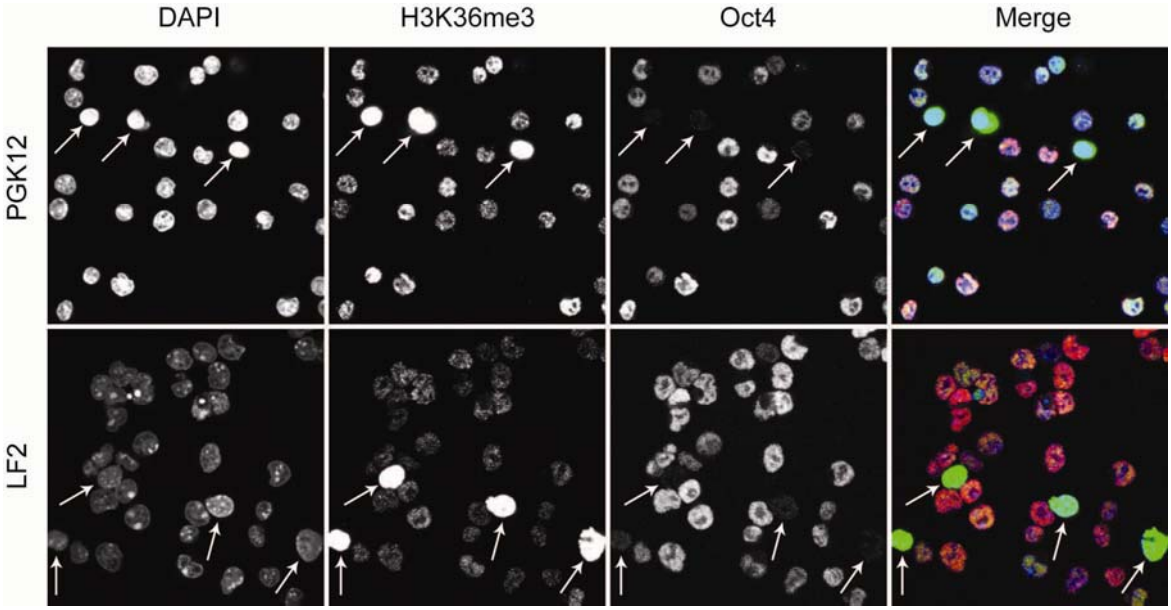

Figure 25. Inverse correlation between H3K36me3 and Oct4 expression in PGK12 and LF2 ES cells

There was a certain population of cells with high level of H3K36me3, but low level of Oct4, for both PGK12 and LF2 cells.

2.2 Identification of H3K36 methyltransferases *in vitro*

Yeast SET2 has been identified as H3K36 methyltransferase *in vivo* (Strahl et al., 2002). Mouse Nsd1 (Rayasam et al., 2003) and human HYPB (Sun et al., 2005) were shown as H3K36 methyltransferases by *in vitro* study. The sequence alignment within SET domain in Figure 26 shows that the conserved residues or motives, required for HMT activity (Dillon et al., 2005), are present in all members of SET2 family HMTs. The alignment predicts that these proteins are putative histone methyltransferases. From the structure features illustrated in Figure 8, SET2 family can be further divided into four groups. Group I includes Ce_Mes4 and Ce_Set12; Group II includes Ce_HypB, Sc_Set2, Dm_HypB and Mm_HypB; Group III includes Dm_Ash1 and Mm_Ash1; Group IV includes Dm_Mes-4, Mm_Nsd1, Mm_Nsd2, and Mm_Nsd3. Nsd1, Nsd2, Nsd3, and HypB will be discussed below.

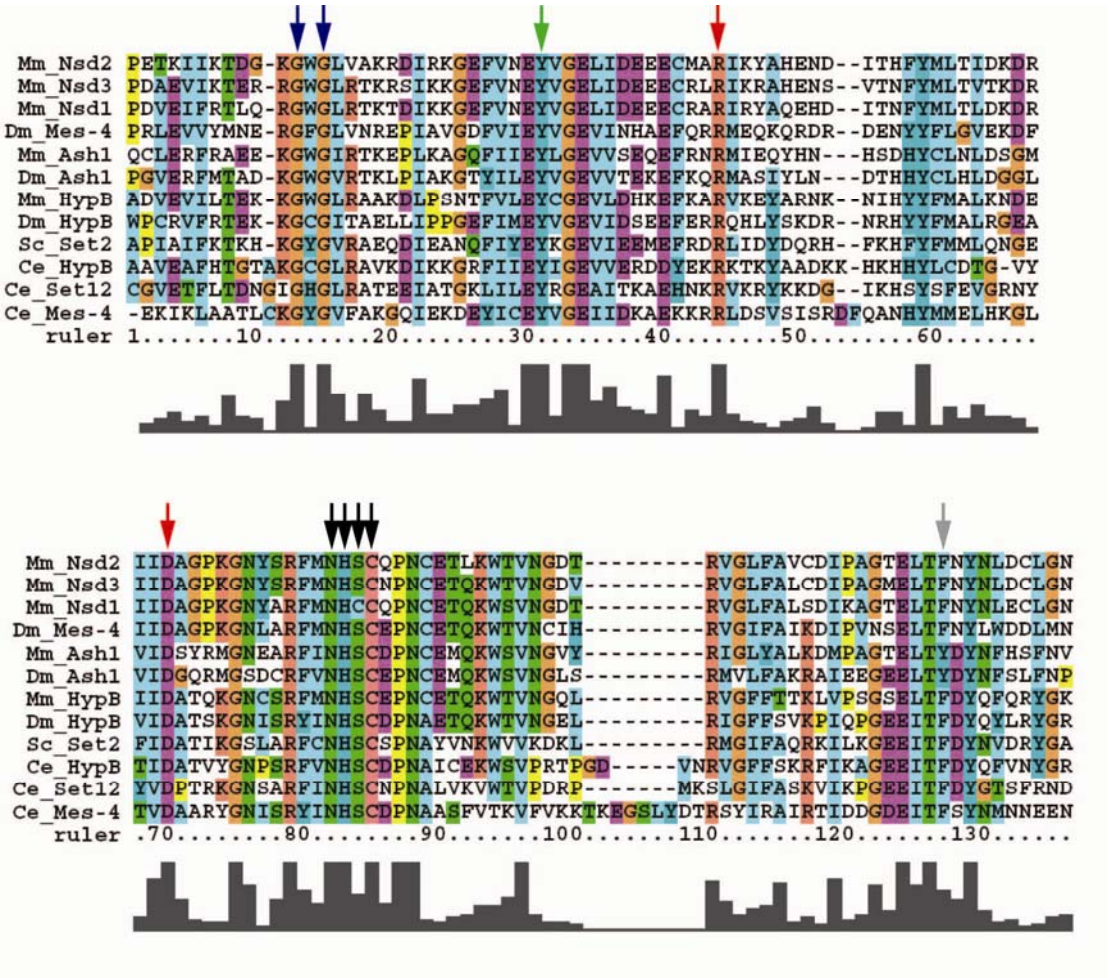

Figure 26. A protein sequence alignment of SET domains of SET2 family HMTs

SET domain is defined by SMART. The alignment was performed using Clustal X. The conserved residues involved in binding to AdoMet and the target lysine (blue arrow), catalysis (green arrow), the structural pseudoknot (black arrow) , an intra-molecular interacting salt bridge (red arrow), and a F/Y switch (grey arrow) controlling whether the product is mono-, di-, or trimethylated histone are indicated. Ce: *Caenorhabditis elegans*; Sc: *Saccharomyces cerevisiae*; Dm: *Drosophila melanogaster*; Mm: *Mus musculus*.

2.2.1 *In vitro* HMT assay of Nsd proteins

Nsd family and HypB proteins were chosen for further analysis. To identify if they have histone methyltransferase activity, *in vitro* HMT assay was performed. The catalytic domain of the candidate enzymes including pre-SET, SET, and post-SET domain were cloned into GST constructs (Figure 27a). Recombinant proteins were expressed in bacteria and purified for the enzymatic reactions (Figure 27b).

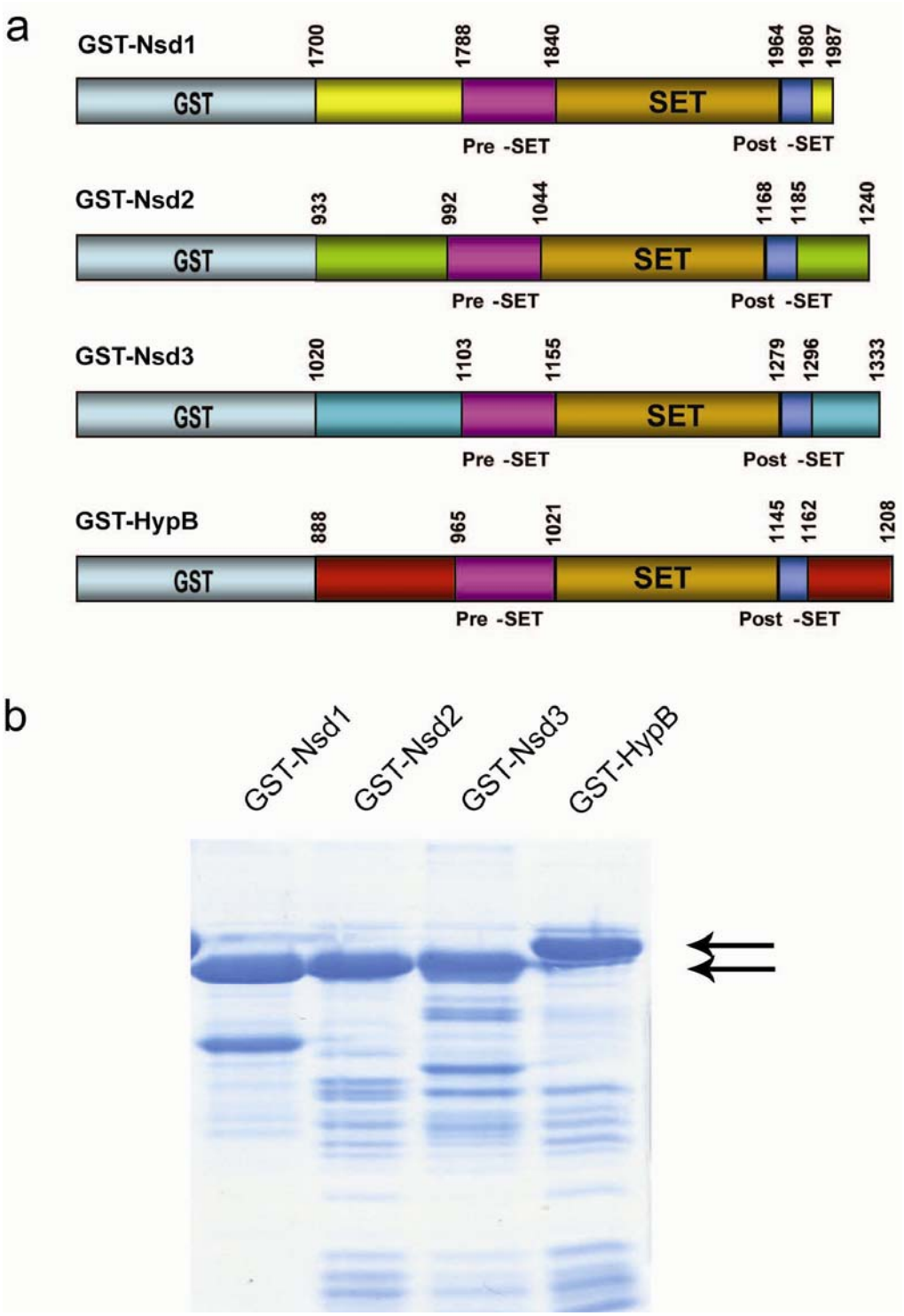

Figure 27. Expression of recombinant GST fusion proteins

(a) Schematic diagram of recombinant GST fusion proteins of putative H3K36 HMTs. Catalytic domain including pre-SET, SET, and post-SET domain, Nsd1 (1700-1987), Nsd2 (933-1240), Nsd3 (1020-1333), and HypB (888-1208), were cloned into GST construct pGEX-6P-1. (b) SDS-PAGE analysis of GST fusion proteins (arrows), visualized by Coomassie staining.

A mixture of histones purified from calf thymus was used as substrate. ^{3}H-labeled S-adenosyl-L-methionine was used as the methyl donor. HMT assay was first performed at pH8.5 and without magnesium condition. The preliminary experiment revealed weak activity from Nsd1, Nsd2, and Nsd3, but strong activity from HypB towards histone H3. It is known that pH and magnesium are critical determinants of the HMT reaction. pH and magnesium concentration were titrated. The titration result confirmed that pH9 was the best for Nsd1 (Rayasam et al., 2003) (Figure 28a). The same pH was used for HMT assay of Nsd2 and Nsd3. The concentration of magnesium was tested at pH9 for Nsd1, Nsd2, and Nsd3. Nsd1 showed moderate activity towards histone H3. Nsd2 and Nsd3 showed weak activity towards histone H3. The activity of Nsd1 is independent on magnesium whereas 10 mM and 5 mM of magnesium are required for the activity of Nsd2 and Nsd3 respectively (Figure 28b).

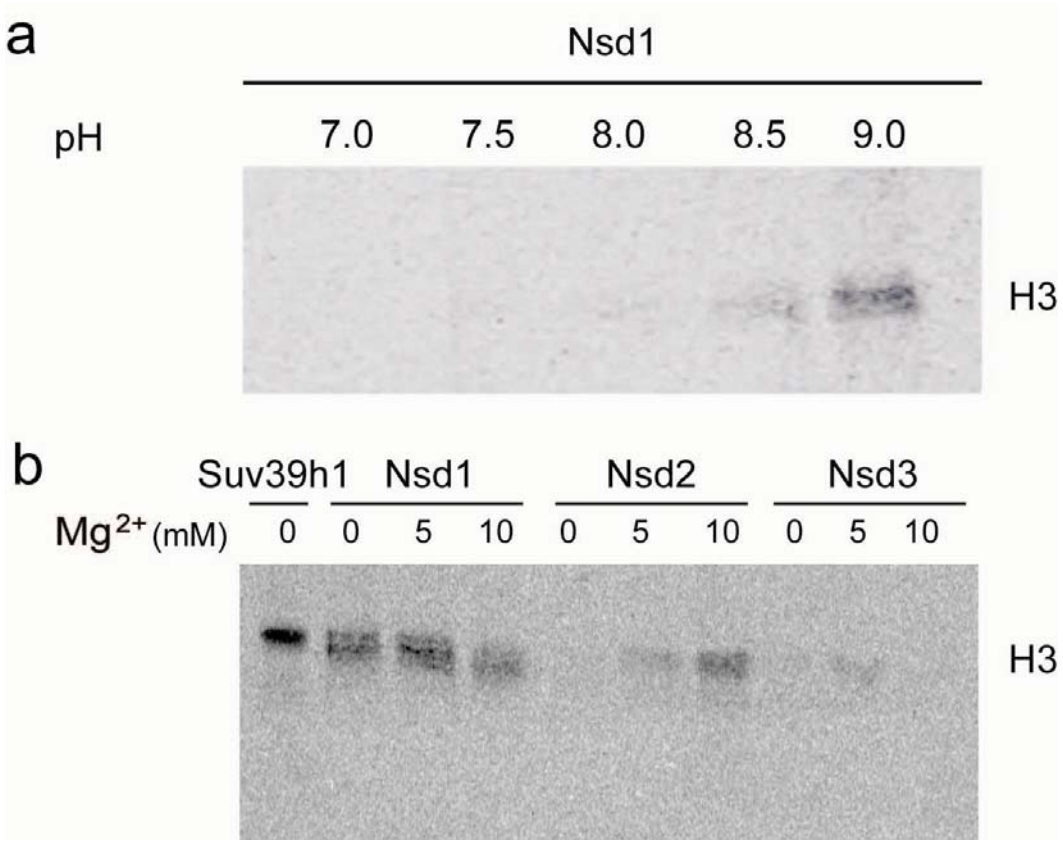

Figure 28. *In vitro* HMT assay of Nsd proteins

(a) pH was titrated for Nsd1 in HMT assay using native histones as substrates for the catalytic domain of Nsd1 proteins. (b) Mg^{2+} concentrations were titrated for Nsd1, Nsd2, and Nsd3. Mg^{2+} is not required for Nsd1 activity. 10 mM and 5 mM of magnesium are required for the activity of Nsd2 and Nsd3 respectively. The H3K9me3 HMT Suv39h1 was used as positive control for the HMT assay (Rea et al., 2000).

2.2.2 *In vitro* HMT assay of HypB

To determine whether the mouse HypB protein is a histone H3K36 methyltransferase as the human protein (Sun et al., 2005), *in vitro* HMT assay was performed with the recombinant protein of HypB catalytic domain. It showed strong activity towards histone H3 and the signal is stronger at pH8.5 and pH9 than at lower pH (Figure 29).The activity of HypB was independent on magnesium (data not shown).

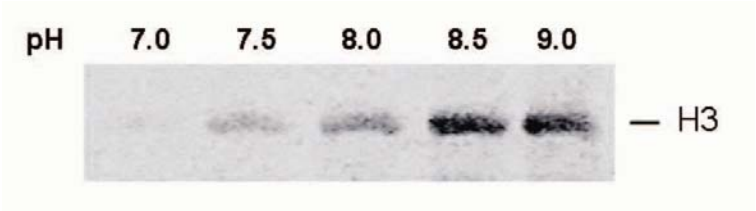

Figure 29. HypB methylated histone H3 *in vitro*

Different pH was tested for the HMT assay of the recombinant catalytic domain of HypB. Native calf thymus histones purchased from Sigma were used as substrates. The recombinant catalytic domain of HypB showed stronger HMT activity at pH8.5 and pH9 than lower pH.

To further identify the targeting residue of HypB, different histone N-terminal peptides were used as substrate in the HMT assay, amino acid numbers are indicated in brackets. HypB showed no activity towards histone H1, H4 (12-31), and H3 (1-20) but H3 (19-38). Mutation at lysine 27 and lysine 37 of H3 (19-38) did not affect the activity of HypB. Only the K36L mutation at lysine 36 of H3 (19-38) abolished the activity of HypB. Thus, H3K36 is the targeting residues of HypB (Figure 30). H3K36 unmodified, monomethylated, and dimethylated but not trimethylated peptides of H3 (19-38) could serve as substrates, suggesting that HypB catalyzes H3K36 mono-, di-, and trimethylation *in vitro* (Figure 30).

HypB could also methylate nucleosomes from wild type and Set2 deleted (Set2Δ) yeast strain, indicating that HypB can methylate nucleosomes with unmodified and pre-methylated H3 lysine 36 residues (Figure 31). Nucleosomes from yeast strains with double deletion of Set2 and Set1 (Set2Δ, Set1Δ) or Set2 and Dot1 (Set2Δ, Dot1Δ), and triple deletion of Set2, Set1, and Dot1 (Set2Δ, Set1Δ, Dot1Δ) also served as substrate for HpyB (Figure 31), indicating HypB can methylate nucleosomes with unmodified H3

lysine 4, 36, and 79 residue. Taken together, these data show that HypB is a H3K36 methyltransferase *in vitro*.

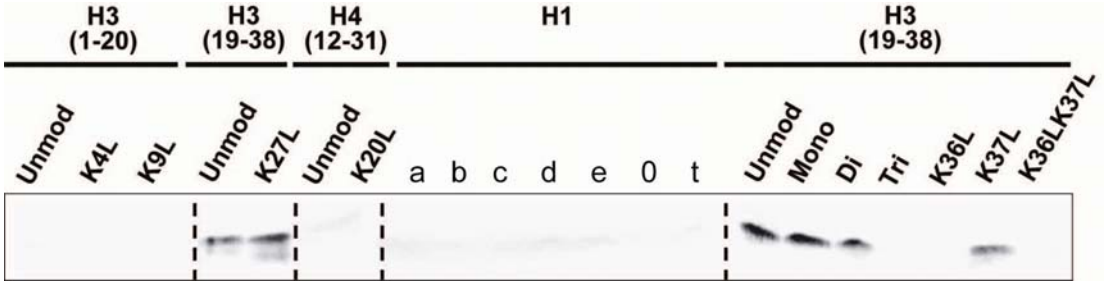

Figure 30. HypB methylated histone H3 peptide at lysine 36

Histone N-terminal peptides were used as substrates in HMT assay to identify the target residue of HypB catalytic domain. The catalytic domain of HypB showed no HMT activity towards H1 peptides of different variants, H4 (12-31), and H3 (1-20). It showed HMT activity towards H3 (19-38). Mutation of lysine 27 and 37 didn't affect the HMT activity. Only mutation of lysine 36 abolished the HMT activity. All H3K36 unmodified, monomethylated, and dimethylated except trimethylated peptides of H3 (19-38) served as substrates.

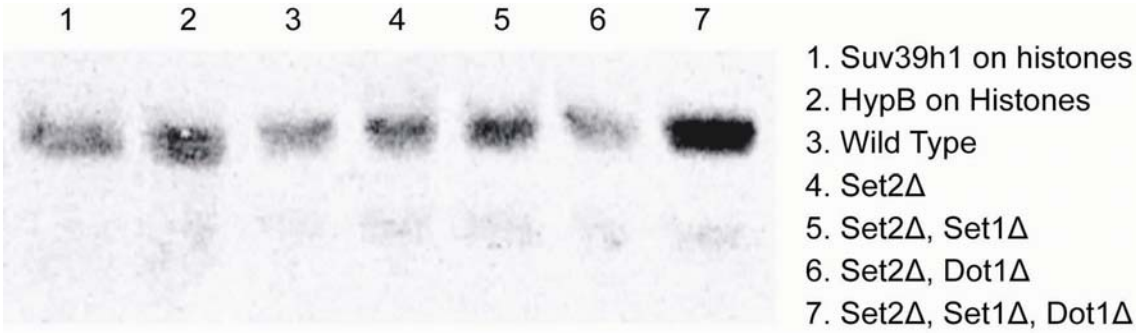

Figure 31. Specificity of HypB on nucleosomes substrates

Specificity of the catalytic domain of HypB was determined using nucleosomes extracted from different yeast strains. The catalytic domain of HypB methylated nucleosomes from yeast strain of wild type, Set2 single deletion (Set2Δ), Set2 and Set1 double deletion (Set2Δ, Set1Δ), Set2 and Dot1 double deletion (Set2Δ, Dot1Δ), and Set2, Set1, Dot1 triple deletion (Set2Δ, Set1Δ, Dot1Δ). Suv39h1 and HypB methylated native histone as positive control of HMT assay.

2.3 Identification of H3K36 methyltransferases *in vivo*

In vitro activity assays were performed with truncated GST-fusion proteins containing only the catalytic domain. To assay the activity of these four candidate enzymes *in vivo*, H3K36 methylation states were analyzed after knockdown by RNA interference of the individual genes in cells by western blot or immunofluorescence (IF) staining using H3K36 specific antibodies.

First, the expression levels of candidate genes were analyzed by quantitative real time PCR in different human and mouse cell lines (Figure 32). The expression levels of NSD1, NSD2, NSD3 and HYPB was higher in human cancer cell lines (293, Hela and U2OS) than in human embryonic carcinoma cell lines (Tera1, Tera2, and NCCIT). NSD2 was highly expressed, NSD1 and HYPB was intermediately expressed, and NSD3 was lowly expressed in 293, Hela, and U2OS cells. Mouse Nsd1, Nsd2, Nsd3 and HypB showed differentially expressed pattern between undifferentiated ES cells (E14, CCE), Trophectoderm Stem cells (TS) and NIH3T3 cells. In ES cells, the expression level of Nsd1 and HypB was higher than that of Nsd2 and Nsd3. In NIH3T3 cells, HypB was highly expressed, Nsd2 and Nsd3 were intermediately expressed, and Nsd1 was lowly expressed. In TS cells, Nsd2 was very highly expressed, HypB was intermediately expressed, and Nsd1 and Nsd3 were lowly expressed.

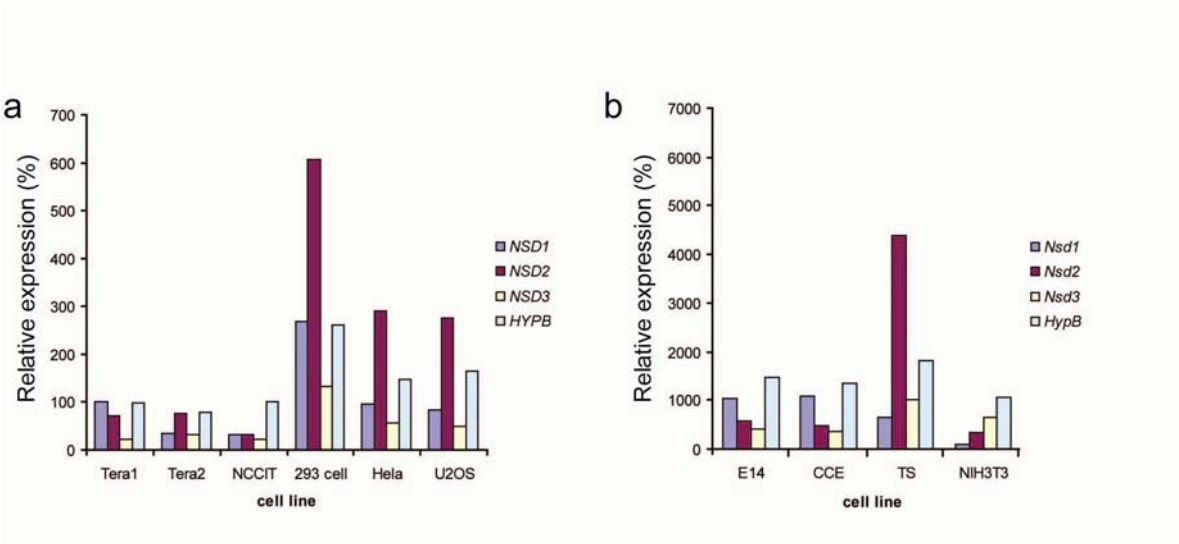

Figure 32. Expression of Nsd and HypB in different human and mouse cell lines

(a) Gene expression was analyzed by real time PCR in human Tera1, Tera2, NCCIT, 293, Hela, and U2OS cells. (b) Gene expression was analyzed by real time PCR in mouse NIH3T3, undifferentiated E14 ES cells, undifferentiated CCE, TS cells. The relative expression was normalized to Nsd1 expression in both human and mouse cells.

Human 293, U2OS and mouse CCE, NIH3T3 were used for further knockdown experiments. The RNAi approach was first set up in 293 cells using shRNA (sigma) (Figure 33). 4 or 5 shRNAs targeting each gene were separately transfected into 293 cells and selected by puromycin for 5 days. The most efficient shRNAs were chosen for further use. 1 shRNA for *NSD1*, 1 shRNA for *NSD2*, combination of 2 shRNA for *NSD3* and combination of 4 shRNA for *HYPB* were transfected into 293 cells and selected for 5 days. mRNA were isolated for gene expression analysis.

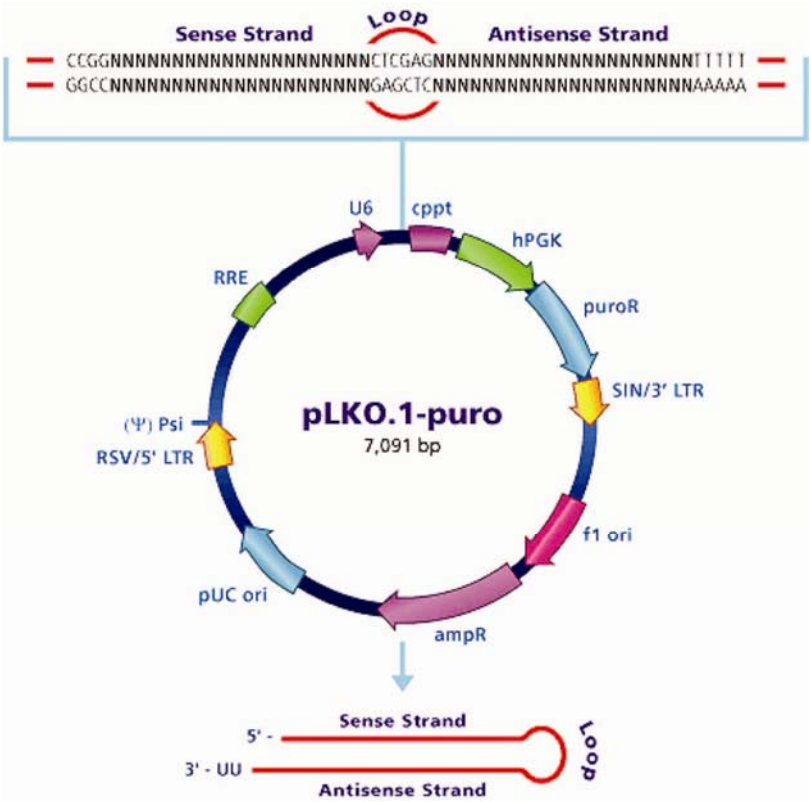

Figure 33 Map of shRNA expressing vector

Transcription of shRNA is driven by U6 promoter. Puromycin resistance gene is used for mammalian selection. Ampicillin resistance gene is used for selection in bacteria.

Efficient knockdown was achieved for all for genes at mRNA level (Figure 34). IF of H3K36 methylation was used to detect the knockdown effect. No change of all three methylation states of H3K36 was observed after knockdown any of *NSD1*, *NSD2*, or *NSD3* gene (data not shown). This may have resulted from redundancy or compensation. Real time PCR analysis showed that the expression of HYPB was upregulated in NSD1 knockdown cells, the expression of NSD1, NSD3, and HYPB were upregulated in NSD2 knockdown cells, and the expression of HypB was upregulated in NSD3 knockdown cells (Figure 34). No efficient knockdown was obtained when all three *NSD* genes were simultaneously knocked down.

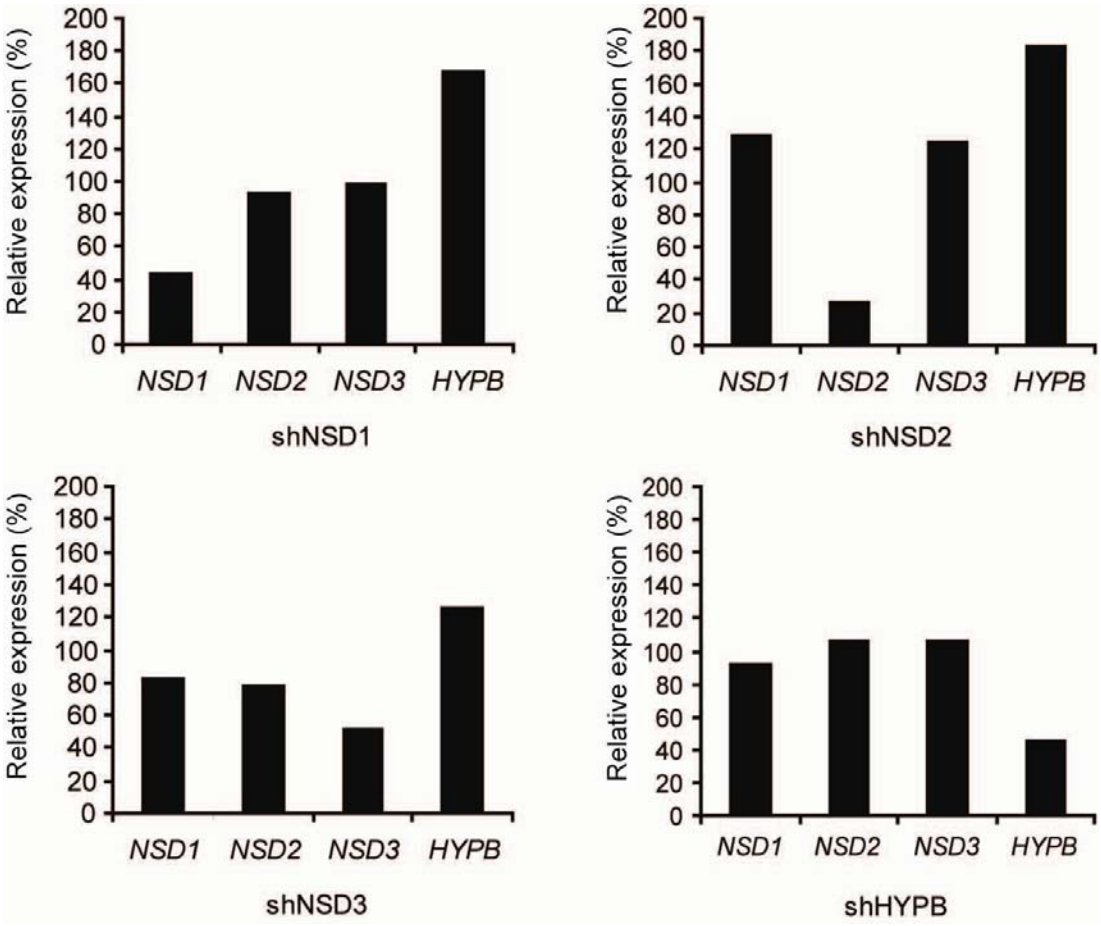

Figure 34. Knockdown of *NSD1*, *NSD2*, *NSD3*, and *HYPB* in 293 cells

Gene expression for all four genes was analyzed by real time PCR in each knockdown sample. Around 60%, 80%, and 55% knockdown was obtained for NSD1, NSD2, and NSD3 respectively. HYPB was upregulated upon NSD1, NSD2 or NSD3 knockdown. Around 60% knockdown was obtained for HYPB, without affecting expression of NSD1, NSD2, and NSD3.

Due to the potential compensation problem for NSD genes, together with the strong activity of HypB *in vitro*, the subsequent study was mainly focused on *HypB* gene. Efficient knockdown of HYPB was achieved in 293 cells using shRNA without apparent changes of *NSD* genes at mRNA level detected by real time PCR (Figure 34). IF data showed that H3K36me3 was strongly reduced in shRNA positive cells indicated by cotransfection of pCX-eGFP construct (Figure 35). In contrast, no apparent change was observed for H3K36me1 and H3K36me2 (data not shown). A small change of H3K36me3 was also detected by western blot (Figure 36). A similar response was observed in human osteosarcoma epithelial cells (U2OS) cells by shRNA (Figure 37) and stealth RNA (Invitrogen) (Figure 38).

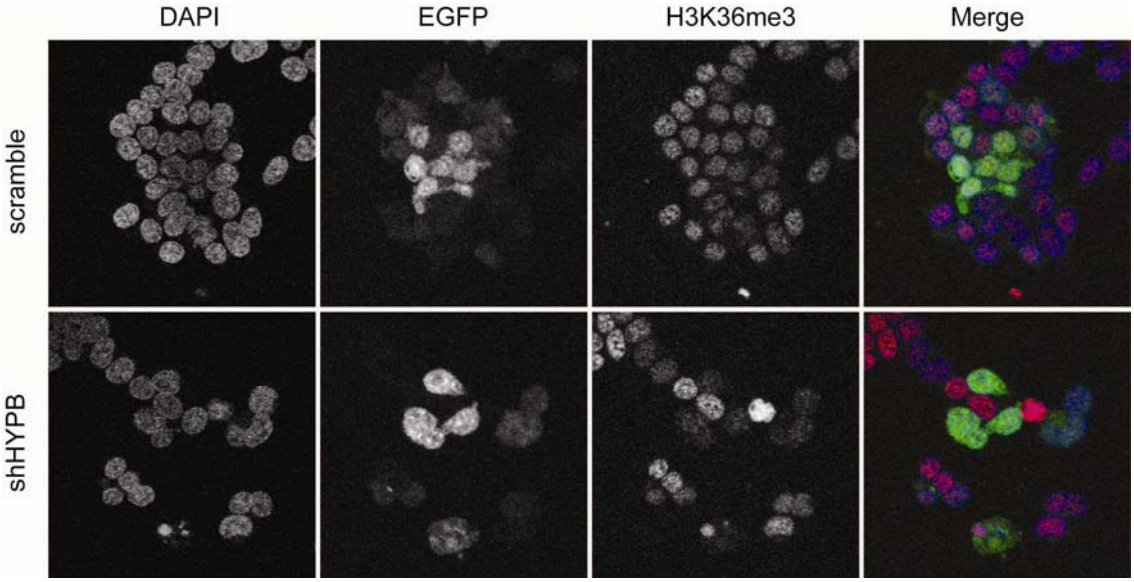

Figure 35. H3K36me3 was strongly reduced upon HYPB knockdown by shRNA in 293 cells detected by IF

shHYPB plasmids were cotransfected with pCX-EGFP plasmid into 293 cells. After 5 days selection by puromycin, H3K36 methylation was detected by IF. H3K36me3 was strongly reduced in shHYPB positive cells indicated by EGFP. No change of H3K36me3 was observed in scramble shRNA transfected cells.

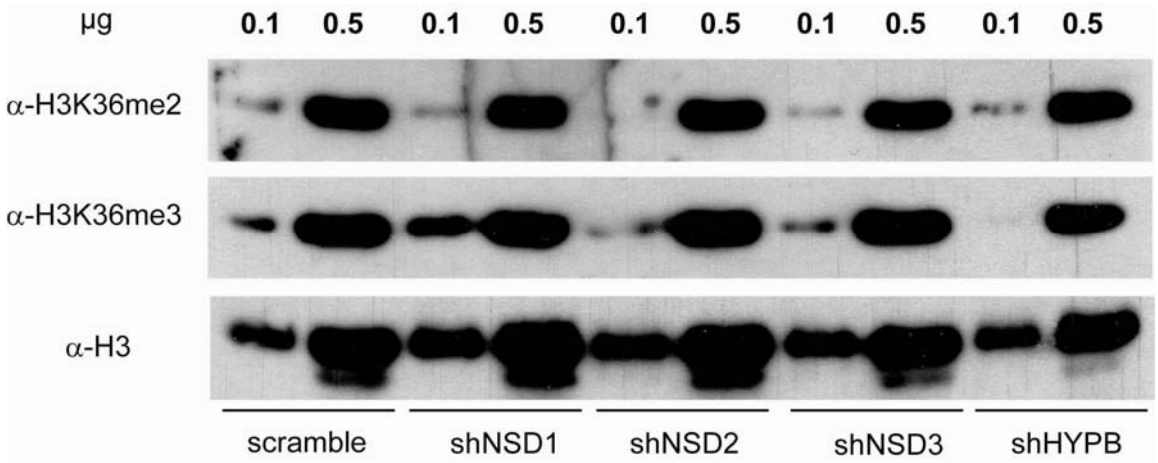

Figure 36. H3K36me3 was reduced upon HYPB knockdown by shRNA in 293 cells detected by western blot

Histones were extracted from cells after 5 days selection with puromycin after shRNA transfection. When 0.1 µg acid extracted histones were loaded, H3K36me3 was clearly reduced in shHYPB samples compared with scramble control sample, but no clear change for H3K36me2. No apparent change of H3K36me2 and H3K36me3 was detected in shNSD1, shNSD2, shNSD3 samples. H3 antibody was used for loading control.

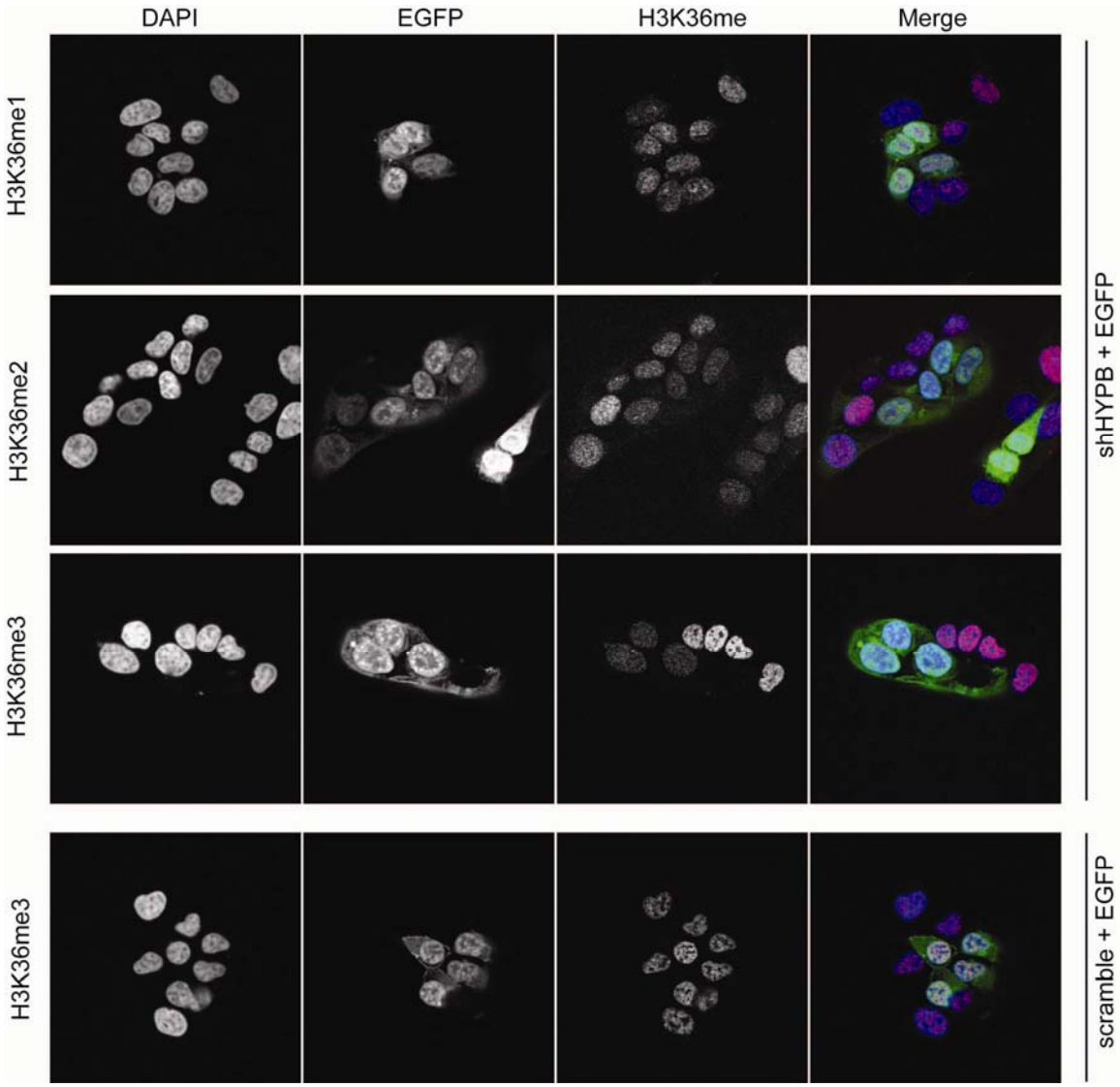

Figure 37. H3K36me3 was strongly reduced upon HYPB knockdown by shRNA in U2OS cells detected by IF

shHYPB plasmid were cotransfected with pCX-EGFP plasmid into U2OS cells. After 3 days without selection, H3K36 methylation was detected by IF. H3K36me3 was significantly reduced, but no apparent change of H3K36me1, and H3K36me2 was observed in shHYPB transfected cells indicated by EGFP. No change of H3K36me3 was observed in scramble shRNA transfected cells.

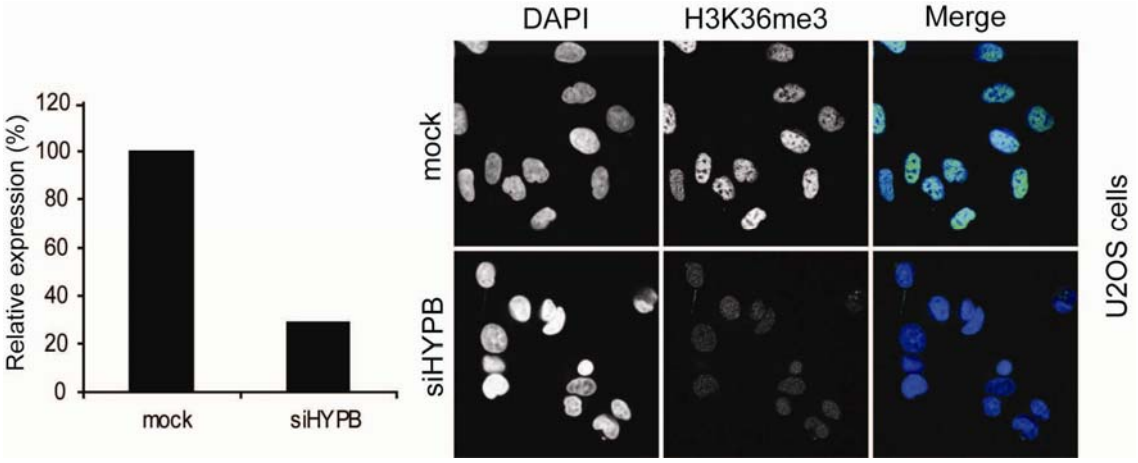

Figure 38. H3K36me3 was strongly reduce upon HYPB knockdown by stealth RNA in U2OS cells detected by IF

Around 70% knockdown was obtained by stealth RNAi detected by real time PCR. IF showed that H3K36me3 was strongly reduced in RNAi sample compared with mock treated control cells.

Subsequently, the function of HypB in mouse system was addressed. Because of unavailability of shRNA for mouse *HypB* gene, stealth RNAs for mouse *HypB* were designed. Stealth RNAs were transiently transfected into NIH3T3 and CCE cells. After 3 days, the knockdown effect was determined by real time PCR of HypB mRNA and IF of H3K36me3. Around 70-80% knockdown could be achieved in NIH3T3 and CCE detected by real time PCR and H3K36me3 was strongly reduced detected by IF (Figure 39). No apparent upregulation of H3K36me1 and H3K36me2 was observed in CCE cells (Figure 40). Thus, HypB is a H3K36 trimethyltransferase *in vivo*.

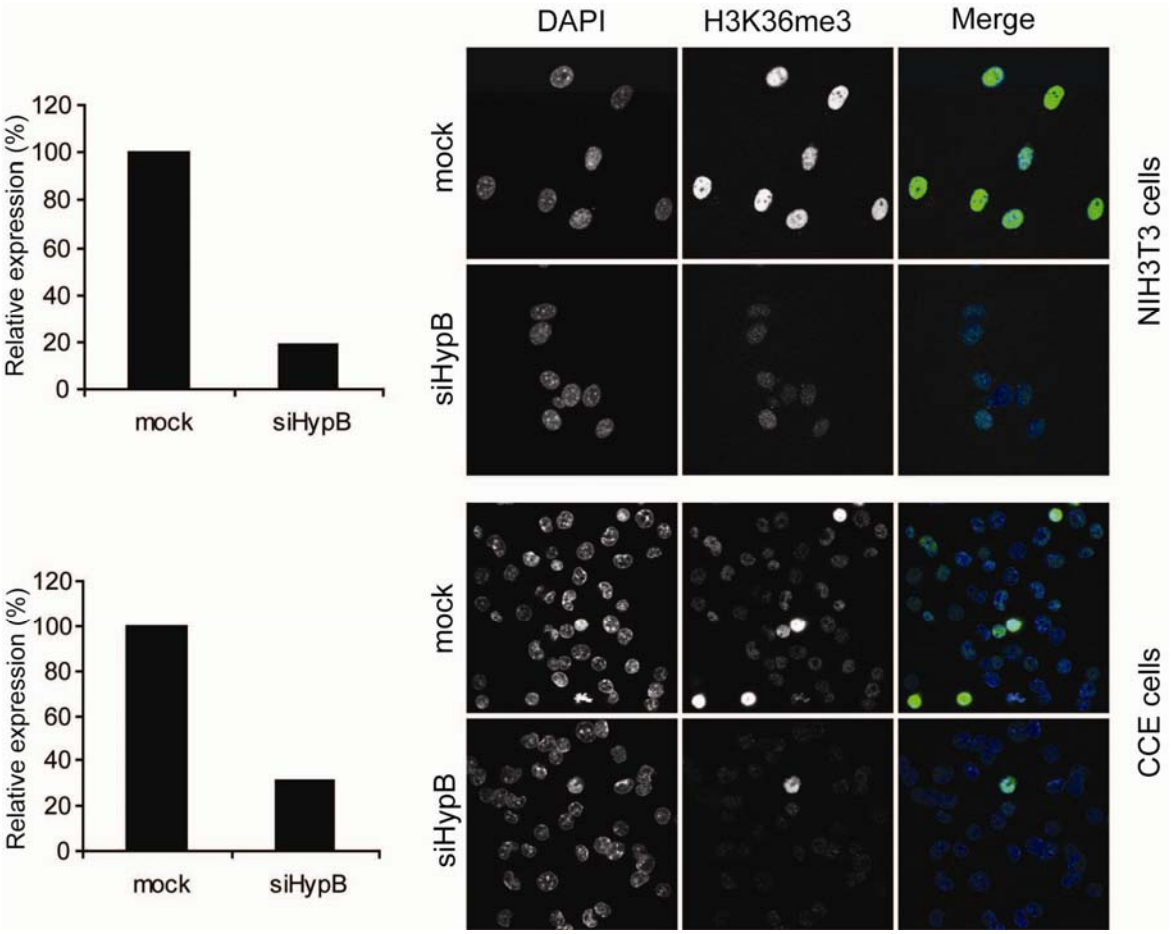

Figure 39. H3K36 was strongly reduced upon HypB knockdown by stealth RNA in NIH3T3 and CCE cells

Around 80% and 70% knockdown was obtained by stealth RNAi after 3 days in NIH3T3 and CCE cells respectively detected by real time PCR. IF showed that H3K36me3 was strongly reduced in RNAi sample compared with mock treated control cells for both NIH3T3 and CCE cells.

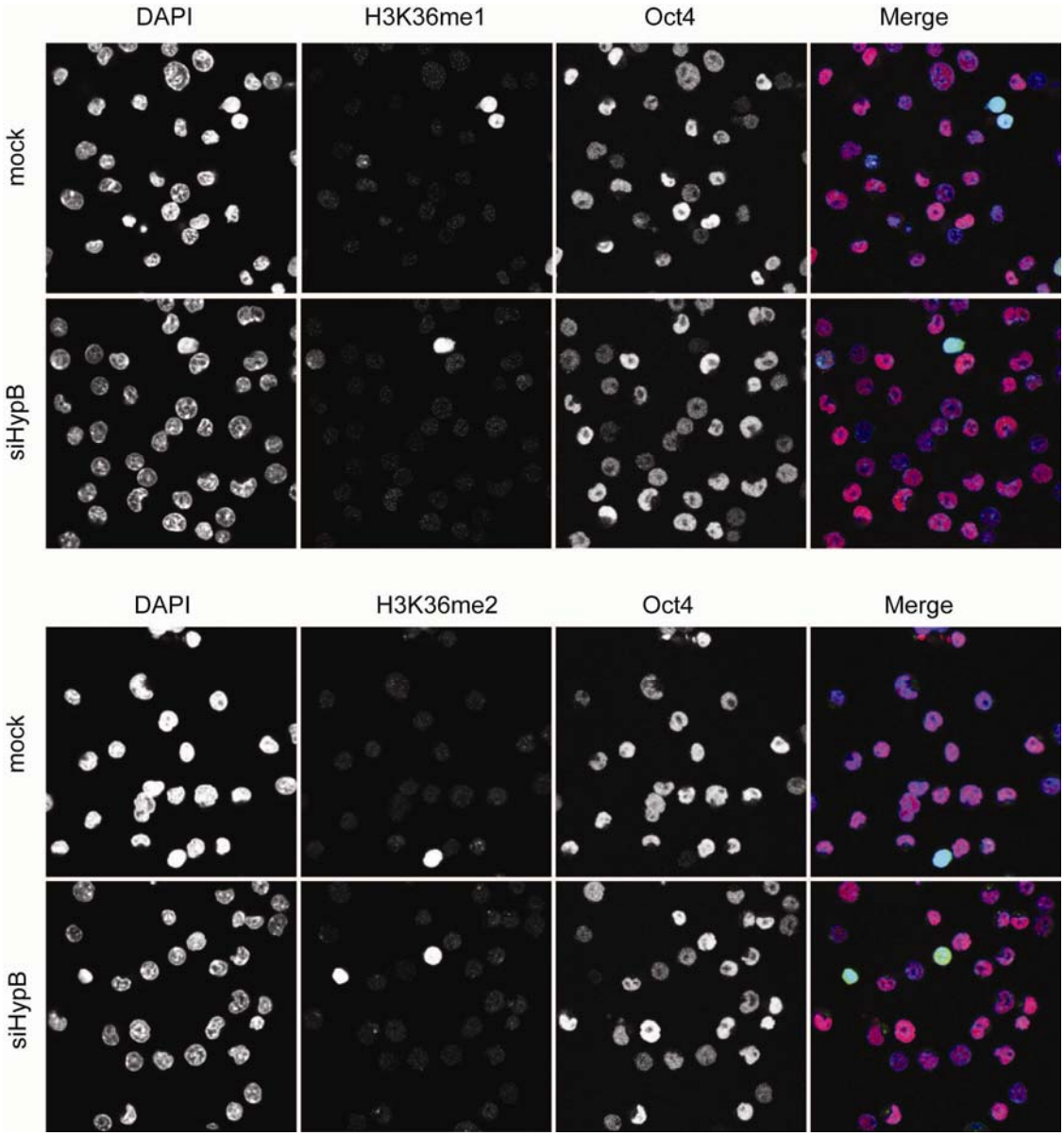

Figure 40. No apparent change of H3K36me1 and H3K36me2 was observed upon HypB knockdown by stealth RNA in CCE cells

No apparent upregulation of H3K36me1 and H3K36me2 was observed upon stealth RNAi of HypB after 3 days in CCE cells.

2.4 HypB facilitates ES cell differentiation

2.4.1 The effect of HypB knockdown in undifferentiated cells

To study the function of H3K36me3 in more detail, the effect of HypB knockdown was monitored in undifferentiated ES cells and during differentiation. First, the effect of HypB knockdown was checked in undifferentiated CCE cells cultured in the presence of LIF. Proliferation rates were analyzed over four days by counting cells. The proliferation rate of HypB knockdown cells is markedly higher than that of mock treated control cells (Figure 41).

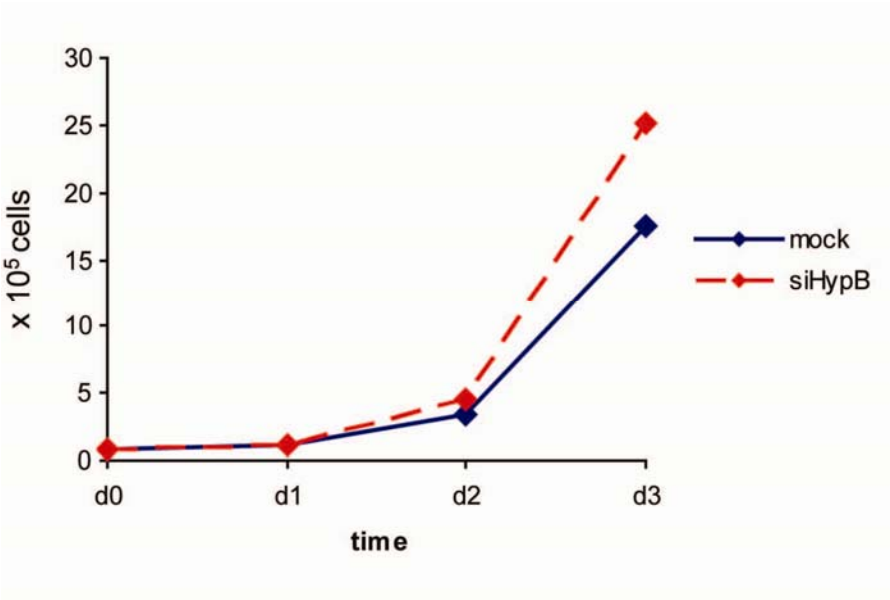

Figure 41. HypB knockdown CCE cells proliferated faster than control cells

Proliferate rate were measured by counting cells, mock treated (blue line) and stealth RNA treated (Red dash line), for four consecutive days after stealth RNA transfection.

To determine the basis of the proliferation difference, the cell cycle profile was analyzed using the BrdU flow kit (BD biosciences). In this method, bromodeoxyuridine (BrdU, an analog of the DNA precursor thymidine) is incorporated into newly synthesized DNA in the S phase of the cell cycle. The incorporated BrdU is detected with specific anti-BrdU fluorescent antibodies by flow cytometry. 7-amino-actinomycin D (7-AAD) was used as

a staining dye for total DNA to define G1, S, or G2/M phases of cell cycle. The cells after 2 days knockdown showed no obvious difference in BrdU incorporation between mock treated and knockdown cells (Figure 42). The distribution of cells during cell cycle is shown in Figure 42.

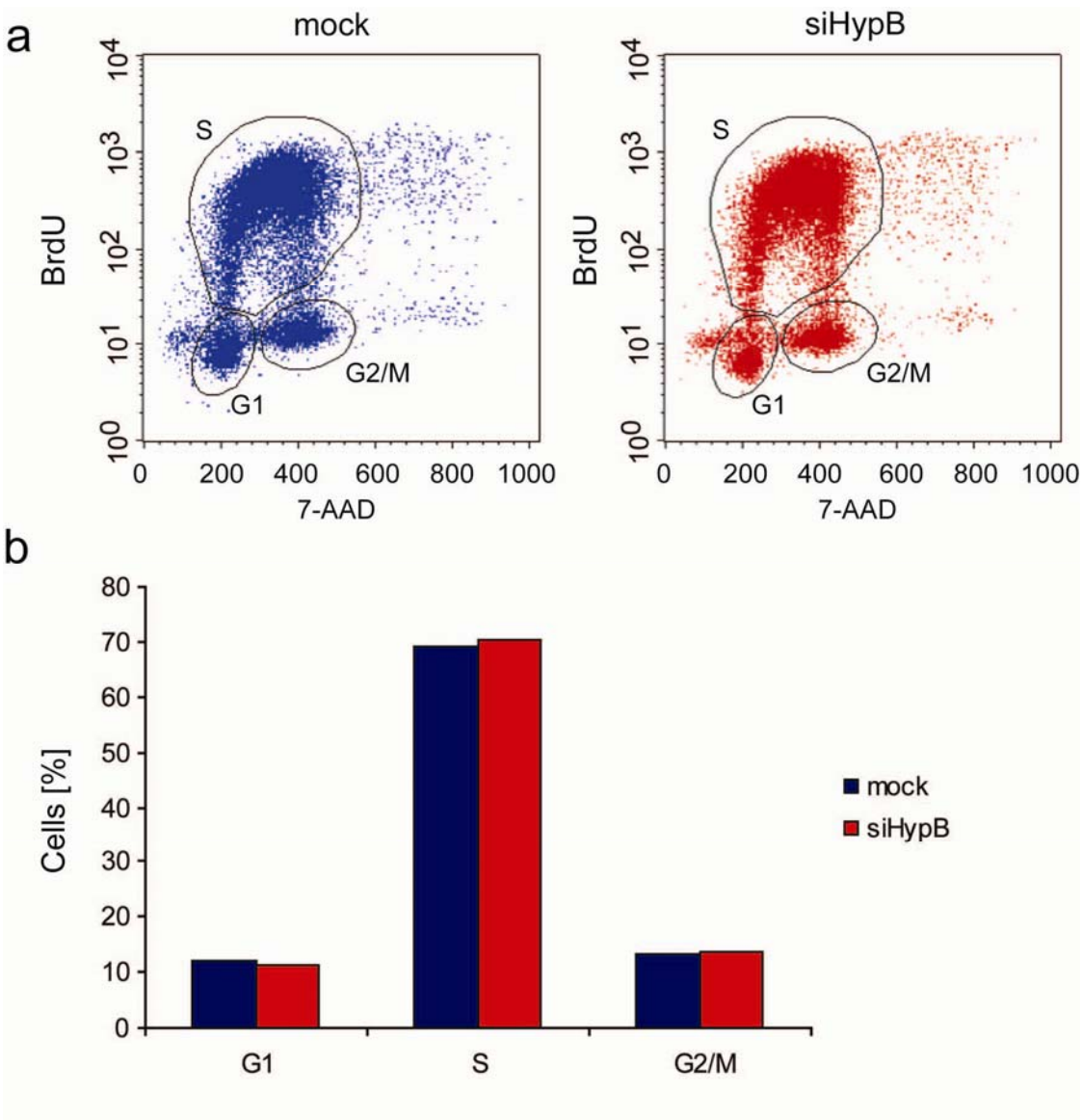

Figure 42. Cell cycle analysis of HypB knockdown CCE ES cells

(a) Flow cytometric analysis of BrdU incorporation in exponentially growing cell. Mock treat: blue; stealth RNA treated: red. (b) Cell cycle distribution. There is no significant difference between mock treated and stealth RNA treated cell. G1: ~11%; S: ~70%; G2/M: ~13%. Total events: ~ 25000.

The decrease of H3K36me3 in HypB knockdown cells was confirmed by IF (Figure 43). The reason for faster proliferating of knockdown cells is not clear. One possibility is that the population of cells undergoing spontaneous differentiation (which are proliferating slower) is smaller in knockdown cells than in mock treated control cells. To test this hypothesis, the relative levels of Oct4 and H3K36me3 as observed by immunofluorescence staining were quantified in single cells. Indeed, the quantification result shows that there were fewer cells with low levels of Oct4 in knockdown cells than in control cells for both day 2 and day 3 (Figure 43).

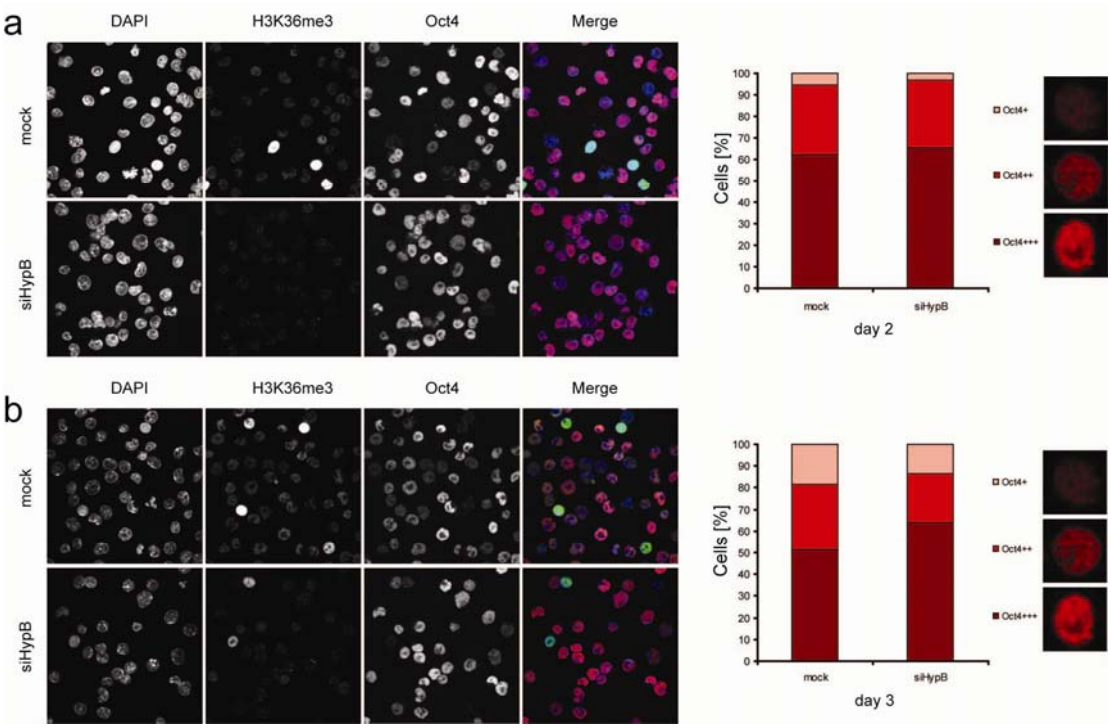

Figure 43. IF of undifferentiated HypB knockdown and control CCE ES cells

H3K36me3 was significantly reduced at day 2 (a) and day 3 (b) upon HypB knockdown. There were slightly less cells with low Oct4 level in the HypB knockdown cells than the mock treated control cells (The category of Oct4 level is described in the following paragraphs).

2.4.2 The effect of HypB knockdown in differentiated cells

An inverse correlation between H3K36 methylation and expression level of Oct4 was observed above. These data are reminiscent of two recent studying on the Mbd3 protein (Gu et al., 2006; Kaji et al., 2007). Mbd3 was found to be required for ES cell differentiation. For example, Oct4 was still expressed in *Mbd3*-null cells after 5 days differentiation induced by LIF removal (Kaji et al., 2006). Mbd3 is a component of the NuRD complex, which also contains the CHD3 protein. CHD3 has been identified as an H3K36me3 binding protein *in vitro* (Mellor, 2006b). It tempting to that H3K36me3 may facilitate the recruitment of Mbd3/NuRD complex to the *Oct4* promoter in the course of differentiation. Thus, an intriguing question is whether knockdown of HypB will show a similar phenotype as *Mbd3*-null ES cells. Thus, the effect of H3K36me3 was subsequently analyzed during differentiation induced by removal of LIF or adding all-trans Retinoic Acid (RA).

The expression level of Oct4 was measured by IF in mock treated and knockdown cells harvested at day 3 at different conditions, undifferentiated (+LIF), differentiated in the absence of LIF (-LIF), and differentiated in the absence of LIF and the presence of RA (-LIF, +RA). The efficiency of knockdown was proven by the decrease of *HypB* mRNA by real time PCR (Figure 44).

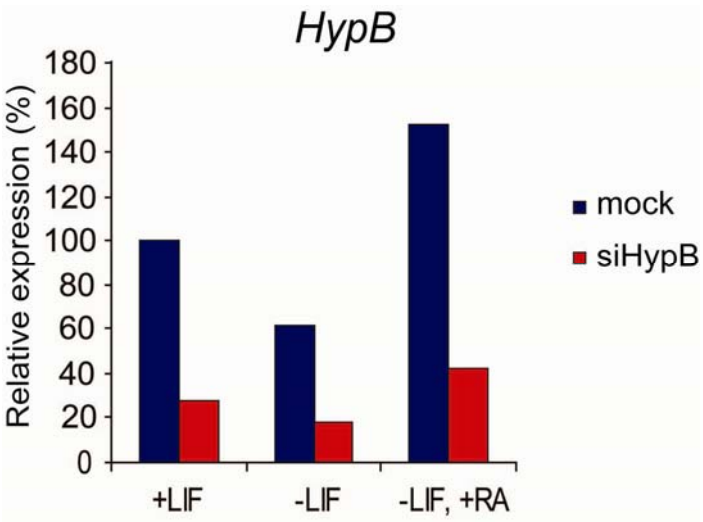

Figure 44. Knockdown of HypB in CCE cells

Around 70% knockdown was obtained for all conditions. The relative expression was normalized to mock control at +LIF condition. In control cells, the level of HypB at -LIF condition is lower than at +LIF condition, in contrast, the level of HypB at –LIF, +RA condition is higher than at +LIF condition.

The decrease in H3K36me3 was detected by IF (Figure 45). These results show that there were clearly more Oct4 positive cells in knockdown cells than mock treated cells after differentiation induced by either -LIF or +RA (Figure 45).

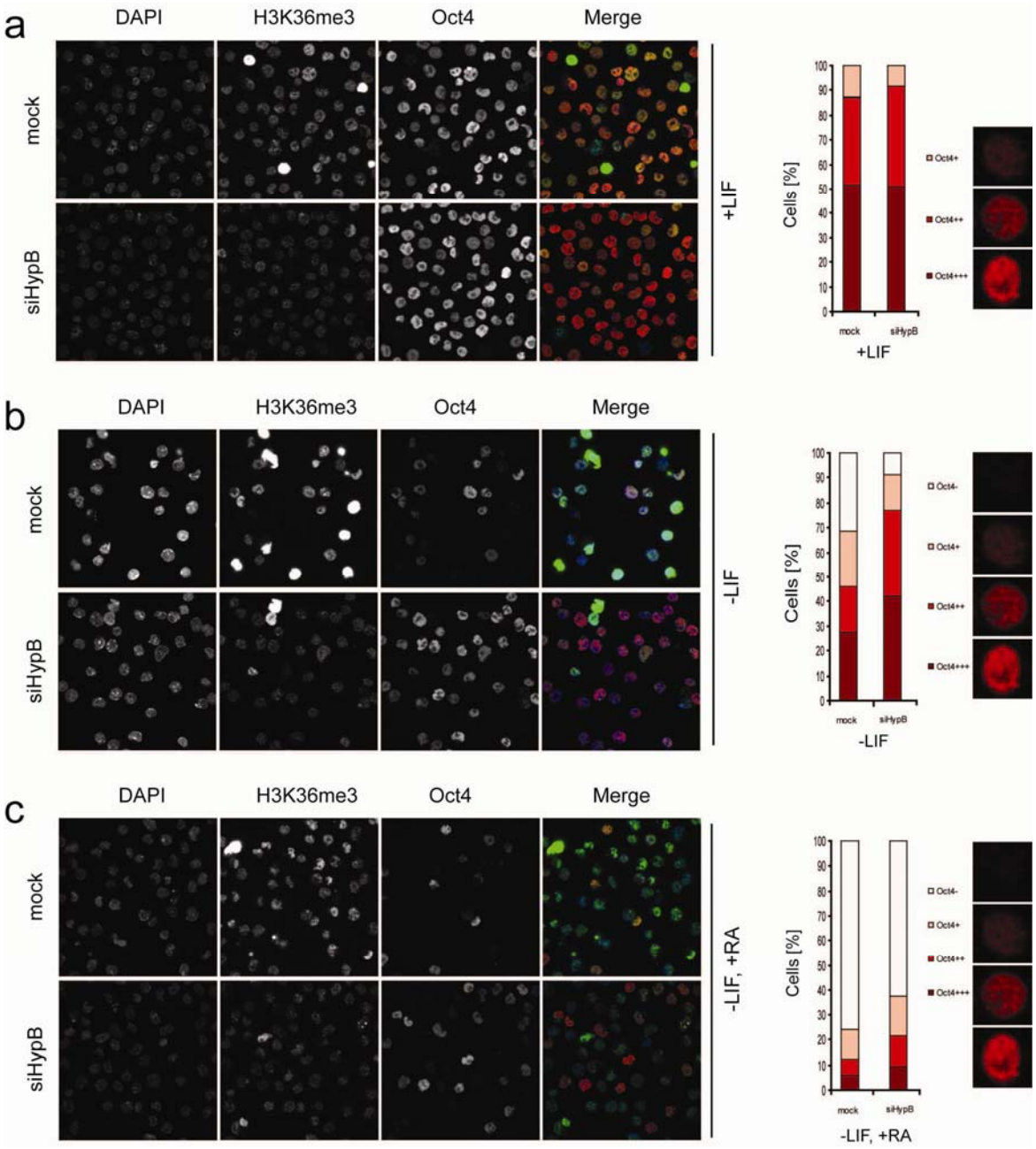

Figure 45. Oct4 repression defects in HypB knockdown ES cells under promotion of differentiation

IF of H3K36me3 showed efficient knockdown for all three conditions. Quantification of IF was shown on the right of the figure. For the mock treated control samples, the number of Oct4 expressing cells is significantly reduced at -LIF than at +LIF (P=3.4e-24). The number is further reduced at –LIF, +RA. There is no significant difference in the percentage of Oct4 expressing cells between knockdown and mock treatment at +LIF (P=0.16). There are significantly more Oct4 expressing cells in the knockdown cells than the mock treated control cells at –LIF (P=2.68e-07) and at –LIF, +RA (P=0.006). Significance was calculated by using the Fisher's exact test (R program).

Further classification of expression level of Oct4 was required for the quantification of IF data. The expression of Oct4 was categorized into four groups based on the intensity of signals, high level (Oct4+++), intermediate level (Oct4++), low level (Oct4+) and no expression (Oct4-) (Figure 45).

The quantification results are shown in Figure 45. For the mock treated control sample, the number of Oct4 expressing cells (Oct4+++, Oct4++, and Oct4+) is significantly less under differentiation induced by LIF removal (-LIF) than undifferentiated cells (+LIF). These data are in agreement with that the expression of Oct4 decreased after the induction of differentiation by removal of LIF (Cartwright et al., 2005). The number of Oct4 expressing cells (Oct4+++, Oct4++, and Oct4+) was further reduced after RA-induced differentiation, confirming the notion that RA is a stronger inducer of differentiation than LIF removal only (Ben-Shushan et al., 1995; Minucci et al., 1996)(Figure 45).

There was no apparent difference in percentage of Oct4 expressing between knockdown and mock treatment under undifferentiated condition (+LIF) (Figure 45a). However, there were significantly more Oct4 expressing cells (Oct4+++ and Oct4++) in the knockdown cells than the mock treated control cells after differentiation, induced by the removal of LIF (-LIF) (Figure 45b). There were more Oct4 expressing cells (Oct4+++, Oct4++, and Oct4+) in the knockdown cells than the mock control cells after differentiation induced by adding RA (-LIF, +RA) (Figure 45c). Thus, in summary, the decrease of the total number of Oct4 expressing cells in each population after differentiation induced by LIF removal (-LIF) or adding RA (-LIF, +RA), is less prominent in the knockdown samples than in the mock treated control samples (Figure 45).

Thus, the IF data shows that the expression of Oct4 failed to be repressed to the same extend in the knockdown cells as the mock treated cells after differentiation induced by either removal of LIF or adding RA, implying that HypB-mediated H3K36me3 functions in differentiating ES cell to facilitate differentiation.

To evaluate the role of HypB in H3K36me3 during differentiation, the level of H3K36me3 was carefully scored in differentiating ES cells. H3K36me3 were categorized into four groups based the intensity of signals: very high (K36me3++++), high (K36me3+++), intermediate (K36me3++), and low (K36me3+) (Figure 46).

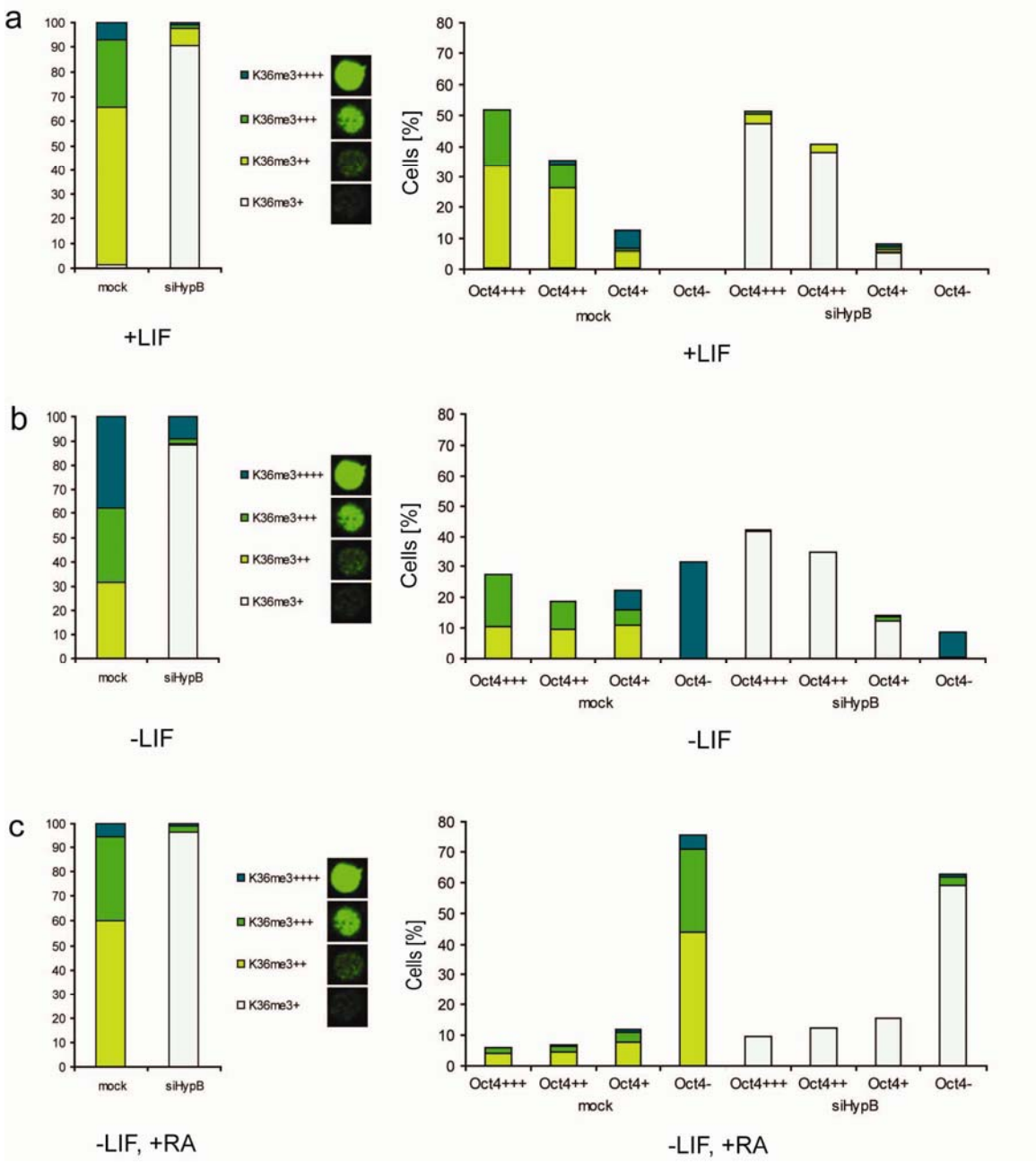

Figure 46. H3K36me3 in mock treated and HypB knockdown ES cells

For the mock treated cells, the number of K36me3+++ and K36me3++ cells significantly increased after differentiation induced by removal of LIF (-LIF) (P=1.32e-14). No apparent difference was observed for the distribution of the H3K36me3 staining classes between undifferentiated (+LIF) and differentiated (-LIF, +RA) condition (P=0.106). The increase of H3Kme3 after LIF removal mainly occurred in the population of Oct4- cells, which suggested that a certain population of cells, with decrease of Oct4 expression gained H3K36me3 upon differentiation. The majority of HypB knockdown cells were H3K36me3+ at all three conditions, indicating efficient knockdown (a, b, and c). Significance was calculated by using the Fisher's exact test (R program).

The distribution of cells in the different groups was first compared for the mock treated control cells. In undifferentiated condition, ~65% cells are K36me3++, ~27% cells are K36me3+++, ~7% cells are K36me3++++, and 1% cells are K36me3+ (Figure 46a). After removing LIF, the number of K36me3+++ cells was significantly increased at the expense of the number of K36me++ cells (Figure 46b). These data indicate that the level of H3K36me3 increased during the early phase of differentiation (see below). A comparison to Oct4 expression levels shows that the cells, strongly stained for H3K36me3, did not express Oct4 (right graphs of Figure 46b), supporting the inverse correlation. In contrast, the distribution of cells with different H3K36me3 levels did not vary between cells differentiated after 3 days RA treatment versus undifferentiated ES cells. Importantly, the majority of cells in the first group were Oct4 negative (Oct4-). Thus, the levels of H3K36me3 in Oct4 negative cells depend greatly on the procedure of *in vitro* differentiation; and thus cell identity of the differentiating cells. It may also reflect a temporal order of differentiation, meaning that H3K36me3 levels first increased during early differentiation (-LIF), followed a dramatic decrease in H3K36me3 induced by RA treatment.

In the HypB knockdown samples, the majority of cells had low levels of H3K36 trimethylation for all three conditions, confirming that the knockdown was effective. The number of Oct4 negative cells with high levels of H3K36me3 was clearly reduced after HypB knockdown when differentiation was induced by removal of LIF. Accordingly, the percentage of Oct4+++ and Oct4++ cells increased. These data suggest that reduced levels of HypB prevent efficient differentiation of ES cells, induced by the absence of LIF. In the presence of RA however, HypB knockdown cells do differentiate, indicating that HypB may only facilitate the differentiation process during the early phase of differentiation.

2.5 Dynamics of H3K36me3 and Oct4 expression during differentiation

To further study the relationship between H3K36me3 and Oct4 expression, the expression of HypB and Oct4, as well as the state of H3K36 methylation during differentiation were monitored by real time PCR and western blot. The differentiation of CCE cells was induced by removal of LIF (-LIF) or adding RA (-LIF, +RA) for 36, 48, 72, 96, and 144 hours. The expression levels of Oct4 only decreased by 25% after the first

96 hours of differentiation induced by LIF removal. In contrast, RA induced a strong and progressive reduction in Oct4 expression (Figure 47a). These data are in agreement with the percentage of Oct4 expressing cells after 3 days of differentiation in the absence of LIF or in the presence of RA (Figure 45b). The dynamic of Oct4 is consistent with current knowledge on Oct4 expression during differentiation.

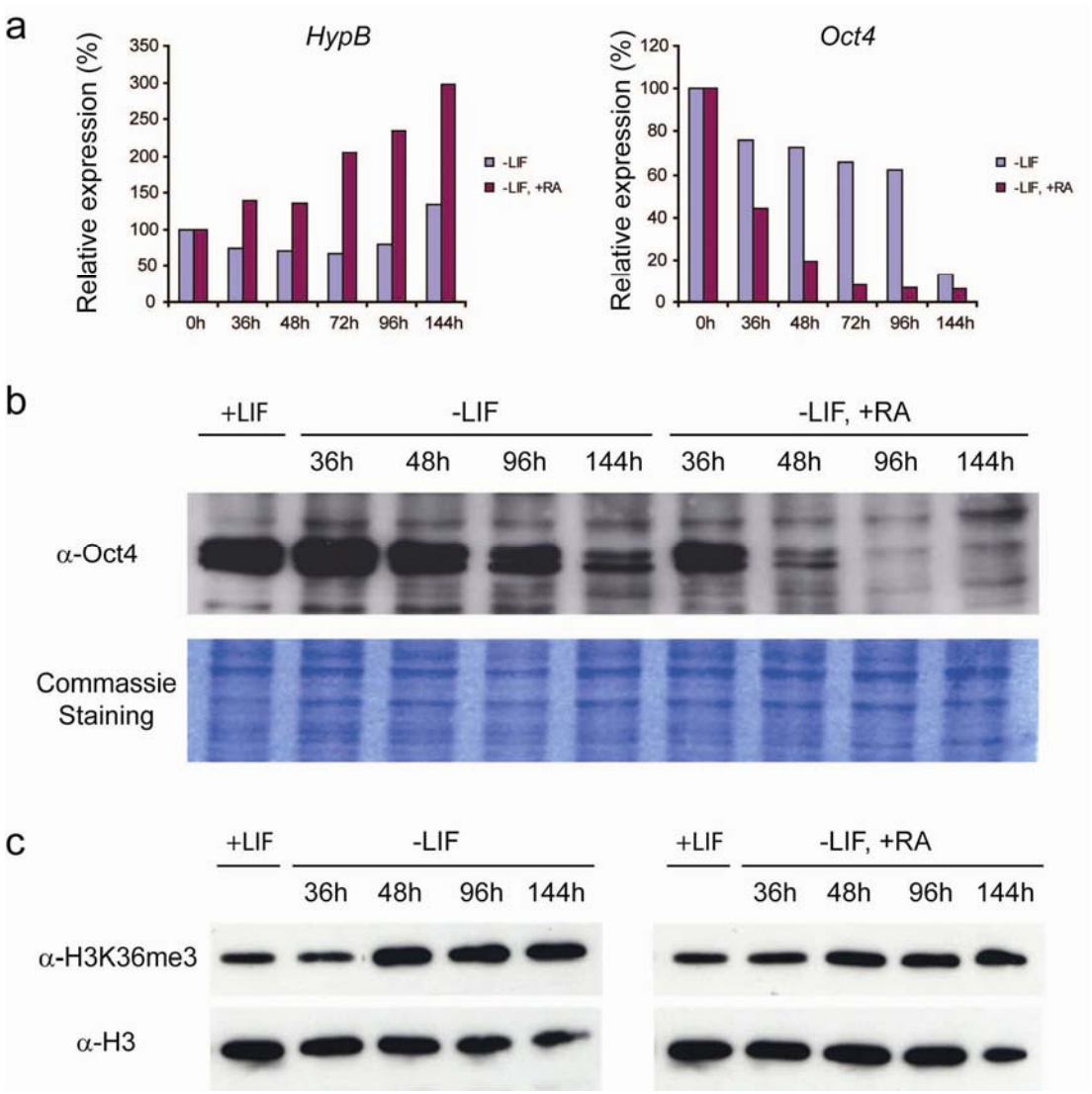

Figure 47. Dynamic of HypB, Oct4, and H3K36me3 during differentiation

Dynamic of Oct4, HypB, and H3K36me3 was analysed during differentiation induced by withdrawing LIF or adding RA at different time. (a) Expression level of Oct4 and HypB was determined by real time PCR. (b) Expression level of Oct4 was determined by western blot. (c) H3K36 trimethylation state was determined by western blot. Histone H3 was used as loading control.

Western blot was performed to determine whether mRNA levels reflect Oct4 protein levels. Cells were fractionated into cytosolic fraction, 0.4M nuclear fraction (nuclear

soluble), and 2M fraction (chromatin bound). Oct4 proteins were mainly present in 0.4M fraction (Figure 47b; data not shown). The kinetics of Oct4 protein disappearance was similar as observed for Oct4 mRNA during differentiation (compare Figure 47b and 47a).

HypB expression slightly decreased at early time (from 36 to 96 hours), but increased at later time (144 hours) after LIF removal. However, HypB expression gradually increased after differentiation induced by RA (Figure 47a).

The methylation states of H3K36me3 were also checked by western blot during differentiation. H3K36me3 increased during differentiation at both conditions of LIF removal and adding RA from 48 hours (Figure 47c). RNAi experiment showed that HypB was responsible for the increase of H3K36me3 during differentiation by LIF removal (Figure 46b). Interestingly, the *HypB* mRNA levels initially dropped down during differentiation in the absence of LIF. It will be important to study the kinetics of HypB protein. Towards this purpose, a HypB-specific antibody has been generated in the lab. Preliminary experiments show that the antibody detects a protein of the predicted molecular weight by western blot. Moreover, it shows a strong nuclear staining immunofluorescence. To explain the increase of H3K36me3, either the HypB protein becomes stabilized and/or becomes enzymatically more activated.

2.6 Gene expression analysis in HypB knockdown ES cells

To investigate the molecular mechanism of the delay in differentiation of HypB knockdown cells, the expression of marker genes during differentiation was analyzed. In these experiments HypB and Oct4 expression showed the same kinetics as observed before (Figure 47a), underscoring the reproducibility the assay during differentiation. The knockdown efficiency was first compared at both undifferentiated and two differentiation conditions at day 3. The knockdown is effective at day 3 for all conditions tested (Figure 48). Independent knockdown experiment lasting for 5 days was also performed. The knockdown efficiency was, however, lower. This may have two reasons: The first is that small interfere RNAs had been diluted out as a consequence of cell division. Alternatively, the increase in HypB expression may have counter-acted the knockdown effect of the siRNA (Figure 48).

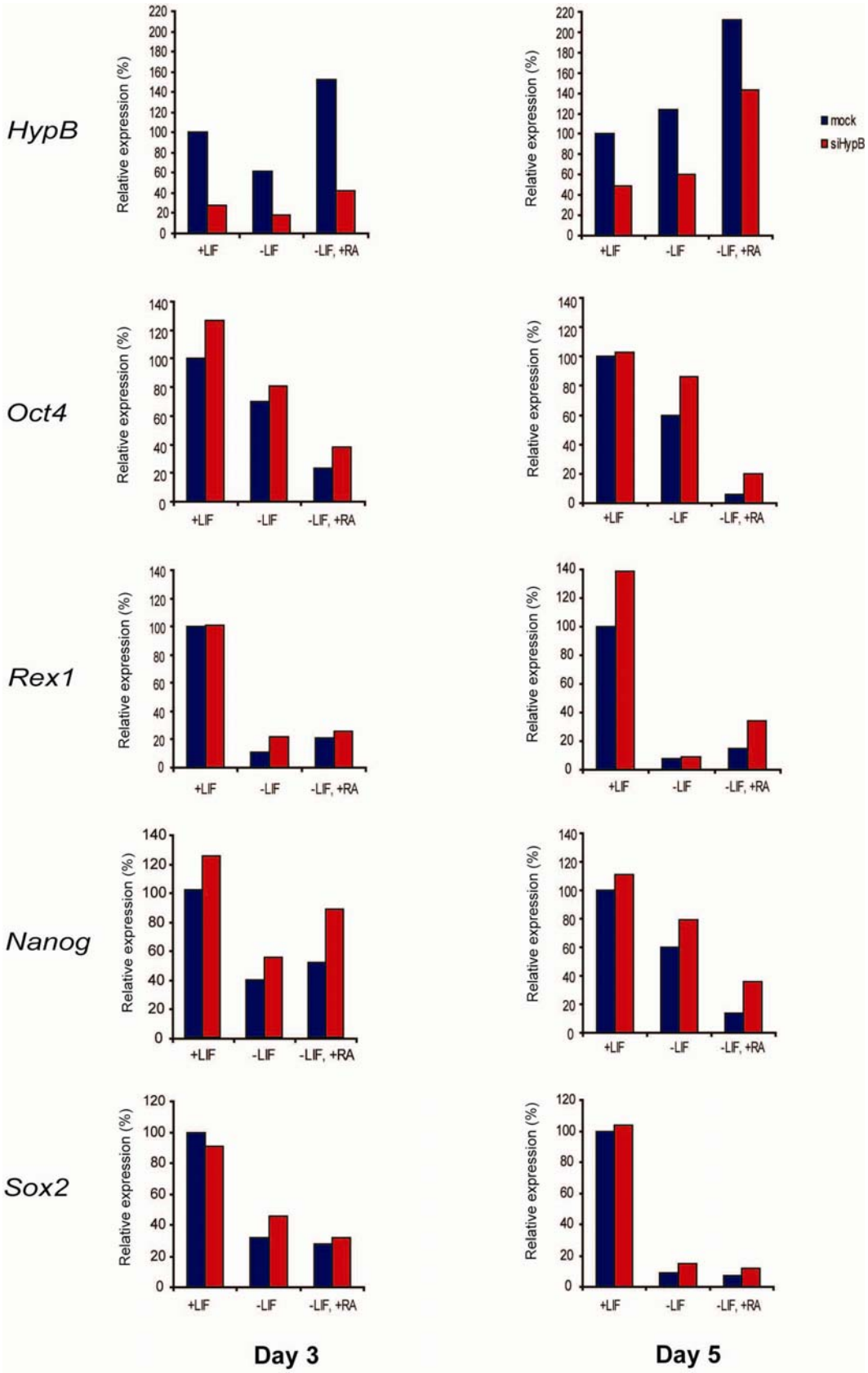

Figure 48. HypB knockdown ES cells fail to repress ES cell markers as control cells upon induction of differentiation

Expression level of *HypB*, *Oct4*, *Rex1*, *Nanog*, and *Sox2* were determined by real time PCR in two independent RNAi experiment with treatment for 3 days and 5 days. The expression levels of each gene in mock cells with LIF were set as 100.

The transcription of genes predominantly expressed in pluripotent cells or in differentiated cells was compared by real time PCR in mock control cells and HypB knockdown cells. The results show that the expression of undifferentiated cell markers *Oct4, Rex1*, *Nanog*, and *Sox2* was slightly higher in knockdown cells than mock treated control cells at day 3 and day 5 after differentiation induced by removal of LIF or adding RA (Figure 48).

In contrast, the expression markers representing specific stages of differentiation (*Fgf5*, *Brachyury*, *Flk1*, *Gata6*, and *Nestin*) were clearly reduced in knockdown cells than mock treated control cells (Figure 49).

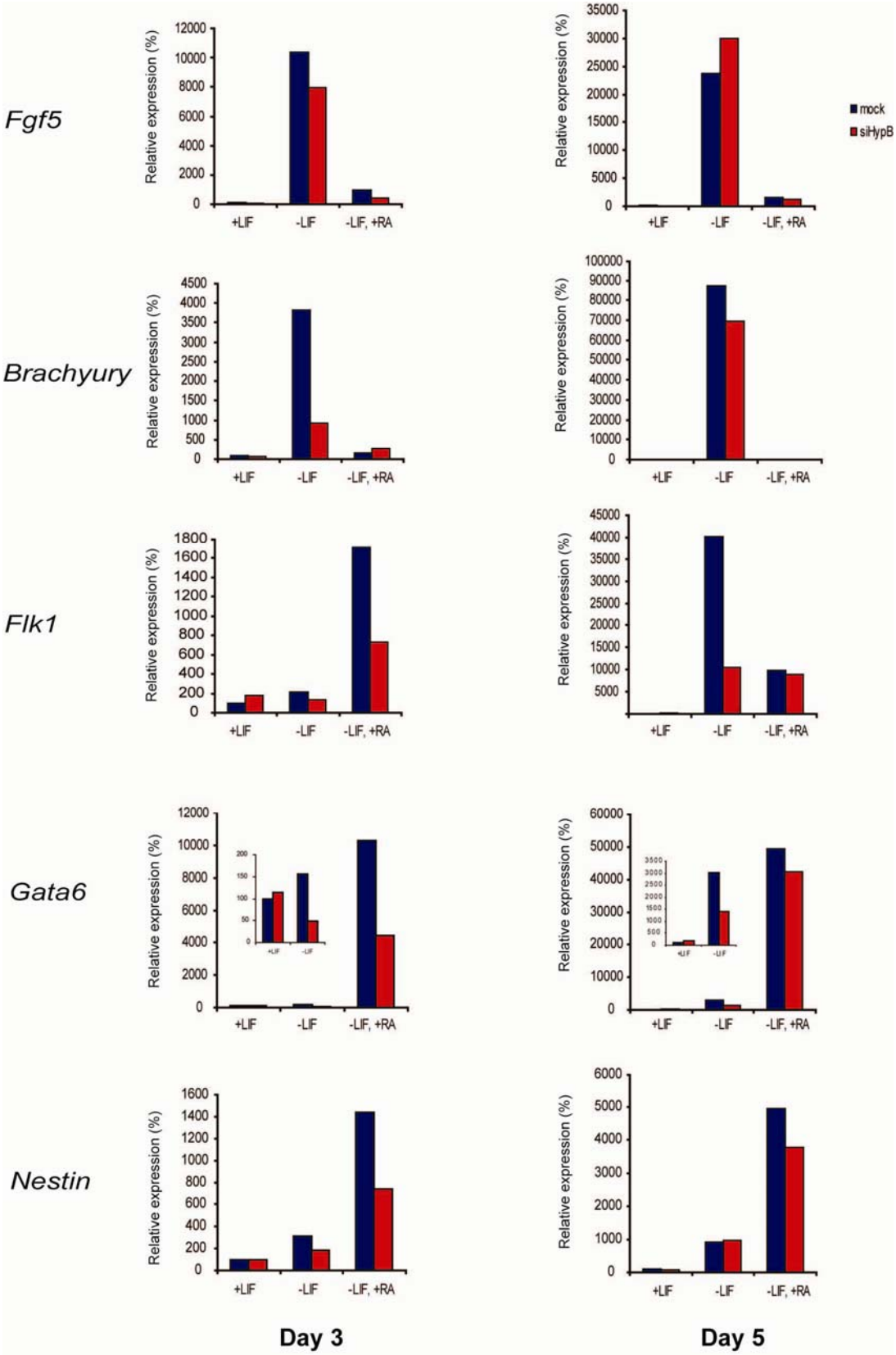

Figure 49. HypB knockdown ES cells fail to activate differentiation-specific genes as control cells upon induction of differentiation

Expression levels of *Brachyury*, *Flk1*, *Fgf5*, *Gata6*, and *Nestin* were determined by real time PCR in two independent RNAi experiment with treatment for 3 days and 5 days. The expression levels of each gene in mock cells with LIF were set as 100.

Fgf5, a primitive ectoderm marker, was barely detectable in undifferentiated ES cells. The levels of *Fgf5* were strongly induced at day 3 and day 5 after differentiation by removal of LIF, but slightly induced at day 3 or day 5 after differentiation by RA (Figure 49). This result is consistent with that *Fgf5* mRNA is transiently upregulated at initial stages of lineage commitment of differentiating ES cells (Cartwright et al., 2005; Hebert et al., 1991; Lowell et al., 2006; Shen and Leder, 1992).

Brachyury and *Flk1*, two mesoderm markers, were not expressed in undifferentiated ES cells. The levels of *Brachyury* were dramatically induced at day 3 and day 5 after differentiation by removal of LIF. The levels of *Brachyury* were low at day 3 and day5 after differentiation induced by RA (Figure 49). This expression of Brachyury is in agreement with that it transiently peaks in early differentiation (Cartwright et al., 2005; Fehling et al., 2003; Ng et al., 2005). The level of *Flk1* was low at day 3, only induced at day 5 after differentiation by removal of LIF and also induced at day 3 after differentiation by RA, then became lower at day 5 (Figure 49). This result reflects the coexpression of *Flk1* and *Brachyury* during differentiation (Huber et al., 2004), but *Flk1* peaks later than Brachyury (Hirst et al., 2006; Kaji et al., 2007; Ng et al., 2005).

Gata6, an endoderm marker, was slightly induced at day 3, but significantly induced at day 5 after differentiation by removal of LIF. The induction of *Gata6* during differentiation was more effective by RA than by removal of LIF, and the level of Gata6 is much higher at day 5 than at day 3 after differentiation by RA (Figure 49). This was confirmed by independent experiment (Figure 50). These results are consistent with the gradually and progressively increase of Gata6 during endoderm differentiation (Hirst et al., 2006).

Nestin, an ectoderm marker, was slightly induced at day 3, but significantly induced at day 5 after differentiation by removal of LIF. The induction of *Nestin* during differentiation was more effective by RA than by removal of LIF, and the level of *Nestin* is much higher at day 5 than at day 3 after differentiation by RA (Figure 49). It showed a similar trend as Gata6 (Figure 50).

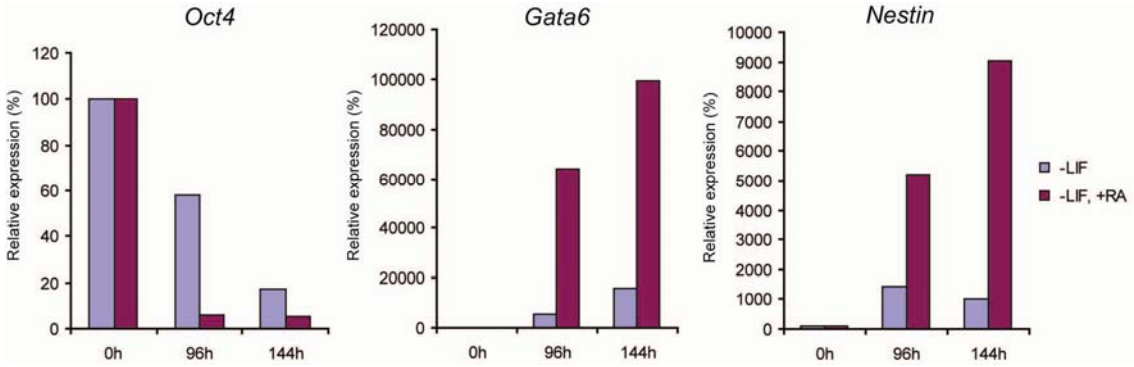

Figure 50. Kinetics of *Gata6* and *Nestin* during differentiation

Gata6 and *Nestin* progressively increased during differentiation particular when induced by RA. The induction of these two genes during differentiation was more effective by RA than by removal of LIF. Oct4 showed similar reduction as before during differentiation.

After HypB knockdown, most of the differentiation specific genes are downregulated. The misregulation of gene expression in HypB knockdown cells is similar as observed in *Mbd3*-null ES cells when differentiation was induced by removal of LIF (Kaji et al., 2006). In addition, *Dppa3*, an early marker of primordial germ cells, was significantly down regulated in HypB knockdown cells (Figure 51); similar as in *Mbd3*-null cells (Kaji et al., 2006). Taken together, these results demonstrate that H3K36me3 by HypB facilitates ES cell differentiation and the knockdown of HypB results in both misregulation of stem cell-specific genes and differentiation-specific genes.

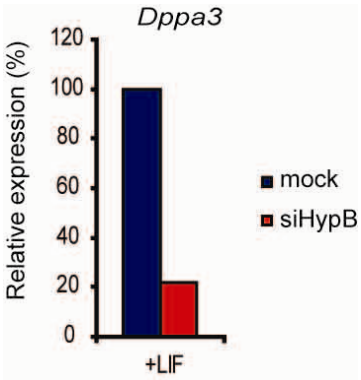

Figure 51. *Dppa3* was downregulated in HypB knockdown ES cells

Dppa3 is downregulated upon HypB knockdown at day 3 in undifferentiated ES cell. *Dppa3* is an early marker of primordial germ cells, which is downregulated in *Mbd3*-null ES cells.

It has been reported that Oct4 failed to be repressed in cells deficient for genes such as *Mbd3*, *Suz12*, *CGBP*, after 5 or 10 days differentiation (Carlone et al., 2005; Kaji et al., 2007; Pasini et al., 2007). To test a possible role of HypB knockdown during long term differentiation, I generated cell line with stable knockdown of HypB by shRNA. The knockdown was efficient (~70%) in undifferentiated cells. However, the efficiency was low after long time of differentiation as observed in RNAi experiment using small interfere RNA (Figure 52). Thus, the analysis of Oct4 expression in 5 or 10 days differentiated cells is unfeasible. *HypB*-null ES cells and mice are required for the further functional analysis of H3K36 methylation during differentiation and development.

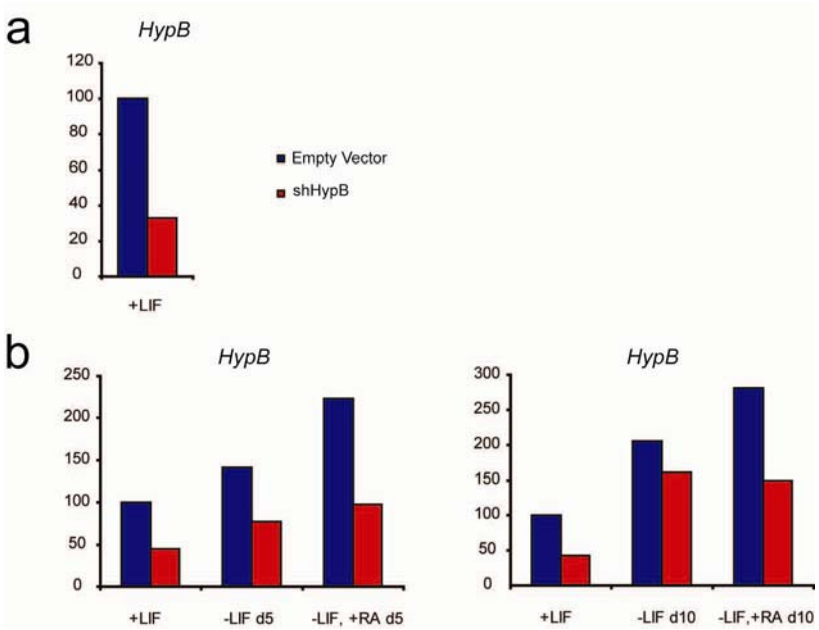

Figure 52. Stable knockdown of HypB in CCE cells

Cell line with stable knockdown of HypB was generated using retrovirus expressing shRNA from pMSCV-LMP vector. Empty vector was used for generating control cell line. In undifferentiated ES cells, the efficiency is from ~55% ~ 70% (a, b). The efficiency become lower at 5 and 10 days differentiated ES cells (b).

2.7 Distribution of H3K36me3 on Oct4 locus during differentiation

To dissect the mechanism of H3K36me3 in regulating Oct4 expression, the distribution of H3K36me3 on Oct4 locus was mapped by chromatin immunoprecipitation (ChIP). In addition to the inverse correlation, several lines of evidences suggested a direct link between H3K36me3 and expression of Oct4. First, as discussed above, there exists a potential link between H3K36me3 and Mbd3/NuRD complex (Mellor, 2006b). Along these lines, it has been shown that GCNF sequentially recruits Mbd3 and Mbd2 to the promoter of *Oct4* and was required for the repression of *Oct4* gene during differentiation induced by RA (Gu et al., 2006) (Figure 14).

Secondly, it has been shown that the proximal enhancer (PE) of *Oct4* is required for its repression during differentiation induced by removal of LIF. *Oct4/lacZ* transgene expression was not downregulated when the PE is absent, but was downregulated when PE is present after the transgenic ES clones were induced to form embryo bodies (Ovitt and Scholer, 1998). The transgenic ES clone with PE deletion construct showed similar phenotype of Oct4 expression as *Mbd3*-null and HypB knockdown cells.

So the question here is whether HypB-mediated H3K36me3 is contributing to the recruitment of Mbd3/NuRD complex to the proximal enhancer and the promoter region of *Oct4* gene during differentiation. Because the dynamics of Oct4 expression is different during differentiation induced by removal of LIF and adding RA, ChIP was performed in both differentiated ES cell, as well as undifferentiated ES cells. Primers covering the *Oct4* locus, including the distal enhancer (DE), the proximal enhancer (PE), the proximal promoter (PP), Intron 1 and Exon 5 were chosen from published paper (Chew et al., 2005) or newly designed (Figure 53).

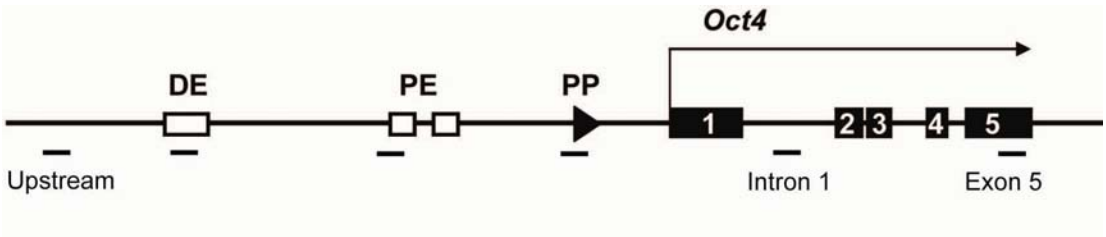

Figure 53. Scheme picture of *Oct4* locus and primers

Start code ATG starts from +1 bp. Upstream: an upstream region of DE (-3155/-2991bp). DE: distal enhancer (-2126/-1956bp). PE: proximal enhancer (-1440/-1191bp). PP: proximal promoter (-377/-194). Intron 1 (+1416/+1620bp). Exon 5 (+4364/+4521bp). ChIP primers are indicated by black bar.

The ChIP data showed that the levels of H3K36me3 were low upstream of transcription start site of *Oct4* (left graph of Figure 54a) and highly enriched in the coding region of *Oct4* in the undifferentiated condition (+LIF). H3K36me3 levels were higher at 3' end (Exon 5) than in the first Intron (right graph of Figure 54a). This result is in agreement with the known distribution pattern of H3K36me3 as a marker for transcription elongation at active locus (Barski et al., 2007; Mikkelsen et al., 2007; Pokholok et al., 2005; Vakoc et al., 2006). The levels of H3K36me3 on coding region (Intron 1 and Exon 5) remained high during differentiation by removal of LIF until 96 hours, and were dramatically reduced after 144 hours (right graph of Figure 54a). This change correlates with the kinetics of Oct4 repression during differentiation (Figure 46a and 46b). In contrast, H3K36me3 on DE and PE was increased 2-3 folds after 96 hours differentiation by removal of LIF (left graph of 54a).

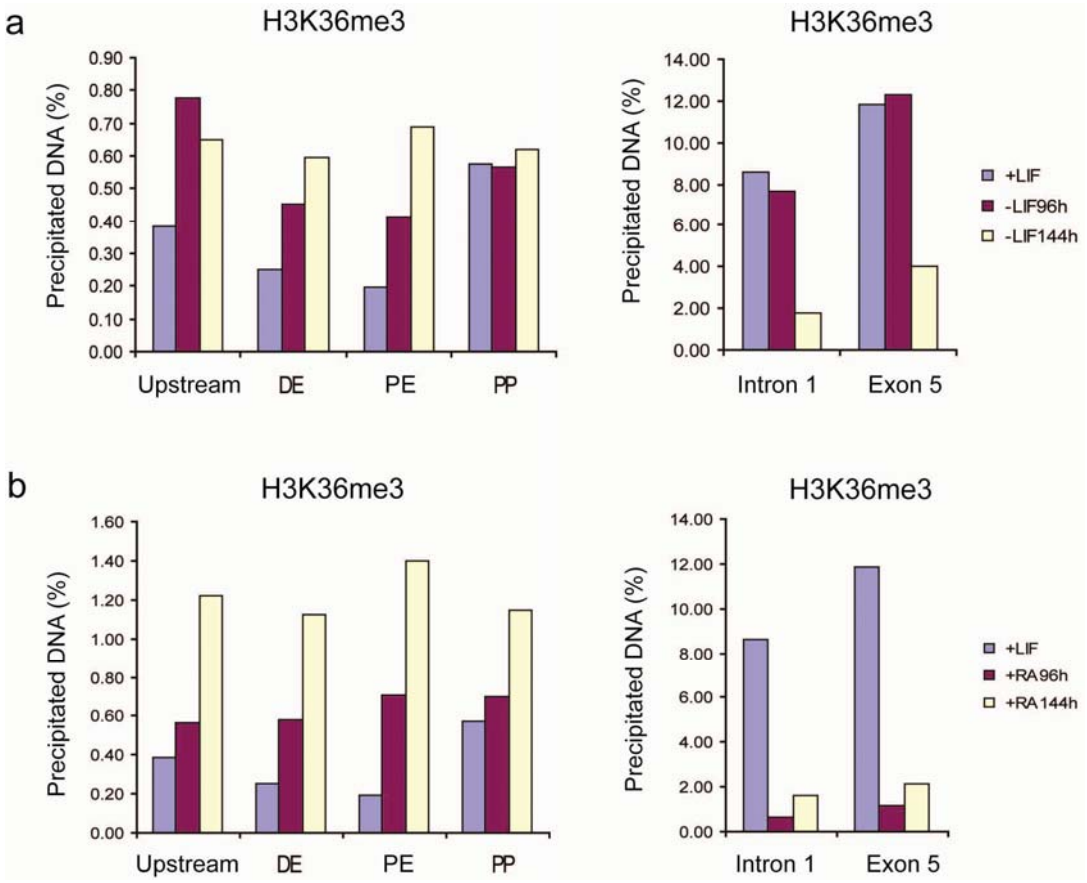

Figure 54. Distribution of H3K36me3 on the *Oct4* locus in undifferentiated and differentiating ES cells

(a) H3K36me3 levels were low upstream of transcription start site and highly enriched in the coding region of *Oct4* at undifferentiated condition (+LIF). After LIF removal, H3K36me3 levels on coding region remained high until 96 hours, and were dramatically reduced after 144 hours. H3K36me3 levels on PE, DE, and upstream of DE increased after 96 hours. (b) H3K36me3 levels on the coding were reduced faster under differentiation by adding RA than by removal of LIF. The levels of H3K36me3 on PP, PE, DE, and upstream of DE all increased after differentiation induced by RA.

The levels of H3K36me3 on the coding region of *Oct4* were dramatically reduced after 96 hours differentiation by adding RA (Figure 54b). The reduction after differentiation is faster under the treatment of RA than under the treatment of LIF removal. These results correlate with the faster reduction of the Oct4 level after differentiation by RA than by LIF removal (Figure 46a and 46b). The level of H3K36me3 increased 2-3 folds upstream of transcription start site (Figure 54b).

The levels of H3K4me3, a marker of active chromatin, was low upstream of the promoter and 3' end of the coding region, but high at the promoter and Intron 1. It is higher at Intron 1 than at promoter in undifferentiated cells (Figure 55a). This result is in agreement with the notion about the distribution of H3K4me3 (Guenther et al., 2007; Mikkelsen et al., 2007; Pokholok et al., 2005; Schneider et al., 2004). The levels of H3K4me3 at the promoter and Intron 1 decreased after 96 hours differentiation by removal of LIF and were reduced to very low level after RA addition (Figure 55a). The low levels of H3K4me3 upstream of the promoter and at 3' end of coding region were further reduced after differentiation (Figure 55a). This change of H3K4me3 also correlated the kinetics of Oct4 during differentiation (Figure 46a and 46b).

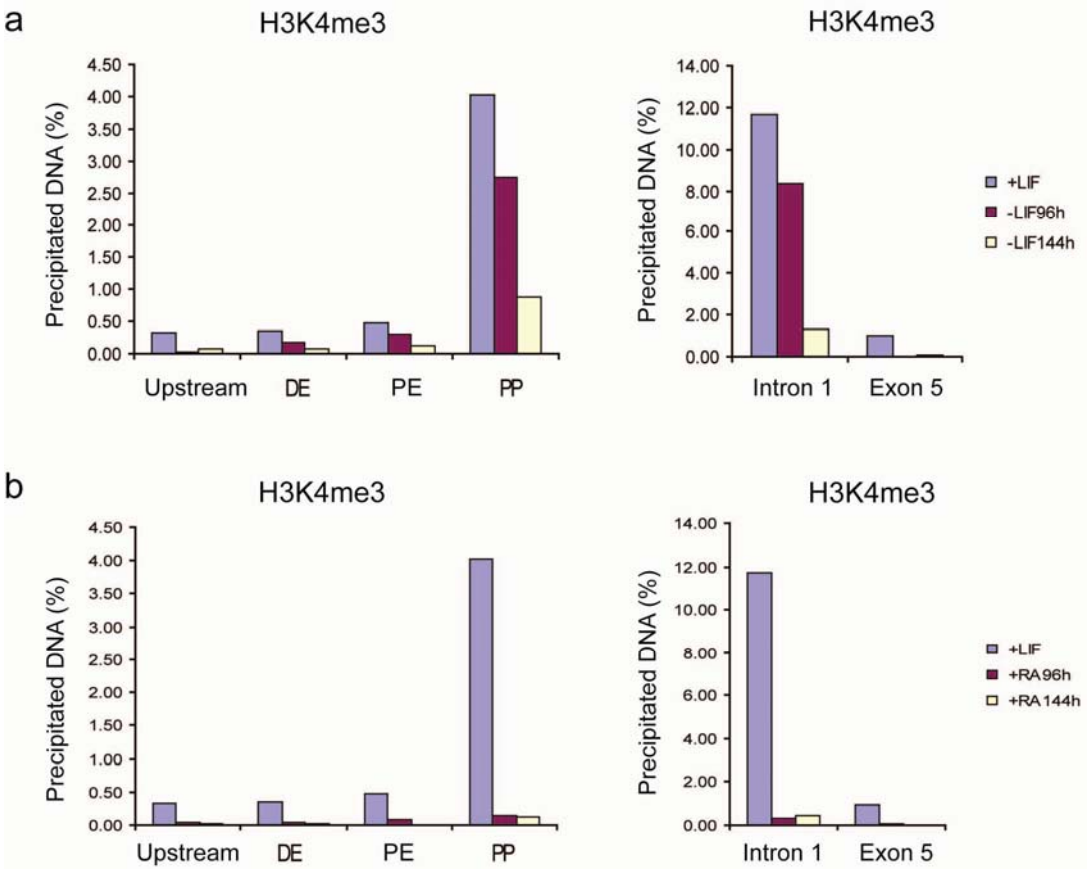

Figure 55. Distribution of H3K4me3 on the *Oct4* locus in undifferentiated and differentiating ES cells

(a) The levels of H3K4me3 were high on PP and Intron 1, but low at upstream, DE, PE and Exon 5 in undifferentiated cells. The levels of H3K4me3 on PP and Intron 1 slightly decreased after 96 hours differentiation but dramatically decreased after 144 hours differentiation by removal of LIF. (b) H3K4me3 decreased to a very low level along the whole Oct4 locus after 96 hours differentiation induced by adding RA.

Taken together, the enrichment of H3K36me3 downstream of transcription start site correlated with the expression of Oct4 during ES cell differentiation. By contrast, the slight enrichment upstream of transcription start site correlated with the repression of Oct4 during ES differentiation.

CHAPTER 3 DISCUSSION

3.1 HypB is the major H3K36me3 methyltransferase in mammalian cells

In yeast, SET2 is the only methyltransferase responsible for all three states of H3K36 methylation (Chu et al., 2006). The deletion of SET2 does not affect cell viability and causes a slight growth defect (Krogan et al., 2003). The major role of SET2-mediated H3K36 methylation is linked to transcription elongation, since H3K36 methylation is enriched in the coding region of genes and this enrichment is dependent on the interaction between SET2 and serine 2 phosphorylated elongating form of RNA polymerase II.

In mammals, several H3K36 methyltransferases have been identified, for example mouse Nsd1 (Rayasam et al., 2003) and human HYPB (Sun et al., 2005). Homolog searching and sequence alignment analysis indicate that more H3K36 methyltransferases exist in mammals. They are structurally different except for the conserved catalytic domain, implicating that different methylation states of H3K36 may be catalyzed by different enzymes that have different cellular function. As deficiency in the *Nsd1* gene causes embryonic lethality (Rayasam et al., 2003), the phenotype is more severe in mammals than in yeast. H3K36 methylation may therefore have more important functions in mammals. Nevertheless, not much is known about H3K36 methylation *in vivo* in mammals.

In this thesis, I identified HypB as a major H3K36 trimethyltransferase *in vivo*. IF shows that H3K36me3 localizes to euchromatin region. HypB knockdown resulted in a significant decrease of H3K36 in both human cells and mouse cells. The domain features HypB and the protein sequence alignment shows that HypB is the closest homolog of yeast SET2. The SRI domain, SET2 Rbp1 Interacting domain, by which SET2 interacts with the elongation RNA polymerase II, is conserved in mammalian HypB (Sun et al., 2005) .

In yeast, an *in vivo* function of H3K36 methylation during transcriptional elongation has been revealed by identifying Eaf3 as a H3K36 methyl binding protein. The H3K36 methyl marker serves to recruit a HDAC (Rpd3) complex through the chromodomain of

Eaf3 and the later identified PHD domain of Rco1 to antagonize H3 and H4 acetylation in the body of genes (Li et al., 2007b). This Set2-Rpd3S pathway controls global acetylation levels at coding region. A subset of genes depends on this pathway to suppress spurious transcriptional initiation at the coding region. In addition, infrequently transcribed and long genes depend on the Set2/Rpd3S pathway for accurate transcription (Li et al., 2007c). It was also reported that Set2 was required for repression of anti-sense transcripts (Nicolas et al., 2007). The average length of higher eukaryote genes is much longer than in yeast. It is conceivable that this pathway would play a more critical role in higher eukaryotes.

In this thesis, chromatin immunoprecipitation in mouse cells showed that H3K36me3 is highly enriched in the coding region of *Oct4* gene in undifferentiated ES cells. H3K36me3 clearly decreases on *Oct4* coding region after its repression after differentiation. These results are consistent that H3K36me3 distributes on the coding region of active genes as a hallmark of elongation. Thus, it will be interesting to investigate whether HypB knockdown will affect the acetylation states of histone acetylation on the coding region and result in inaccurate transcription of genes. Multi exon transcriptome analysis of HypB knockdown cells will allow the identification of spurious transcripts in mammalian system. Consequently, it will be interesting to determine whether such transcripts evoke a transcriptional interference or RNAi-mediated repression response.

The mammalian homolog of yeast Eaf3, MRG15 has been identified as a H3K36 methyl binding protein (Zhang et al., 2006b). MRG15 associates with RBP2 and downregulates the H3K4 methylation at transcribed regions (Hayakawa et al., 2007). It is interesting to speculate that H3K36 methylation on the coding region may serve as marker for the recruitment of MRG15. Another homolog of yeast Eaf3, Msl3l is speculated as a H3K36 methyl binding protein, yet experimental evidence is still lacking. Recently, targeting of MSL3 was found toward the 3' end of the gene and dependent on gene transcription (Kind and Akhtar, 2007). It remains elusive whether this targeting is through H3K36 methylation.

Moreover, being a component of NuRD complex, CHD3 can bind to H3K36me3 through its PHD domain (Mellor, 2006b). This may suggest that the NuRD complex is recruited

to the coding region of transcribed genes as well. It will be interesting to determine whether NuRD complex is target to the coding region of gene through H3K36me3 and has function in controlling gene expression.

The *C. elegans* H3K36 trimethylase (*Met-1*) and H3K9 trimethylase (*Met-2*) affect the activity of each other (Andersen and Horvitz, 2007). Concurrence of H3K36me3 and H3K9me3 on coding region of some genes and existence of a demethylase targeting both H3K36me3 and H3K9me3 in mammals also indicated that there might be interplay between H3K36me3 and H3K9me3. Thus, it is worthwhile to study the methylation state of H3K9 after HypB knockdown to address if H3K36me3 crosstalk with the other histone modifications.

Finally, no apparent change of H3K36me1 and H3K36me2 was detected in HypB knockdown cells, suggesting that these two methyl markers are catalyzed by some other enzymes in separate pathway. It further indicates that H3K36me1 and H3K36me2 do not compensate for the reduced H3K36me3 levels upon HypB knockdown. I cannot, however, exclude that HypB may also function in mediating H3K36me1 and H3K36me2.

3.2 HypB facilitates ES cell differentiation: a role within the coding region of target gene?

Using IF, I observed a clear increase of H3K36me3 in 3 days differentiated ES cells induced by LIF removal. This increase significantly correlated with a reduction in Oct4 expression. Western blot result confirmed the increase of H3K36me3, which also correlated with the decrease of Oct4 expression during ES cell differentiation. Even under the culture condition directing maintenance of the undifferentiated state of ES cells, a small population cells spontaneously differentiates based on the loss of Oct4 expression detected by IF. All three methylation states of H3K36 show an increase signal intensity early differentiating cells, compared to undifferentiated cells with high levels of Oct4. These results indicate that H3K36me3 may have a function during ES cell differentiation.

I found that ES cells differentiation is suppressed upon HypB knockdown when differentiation is induced by removal of LIF or adding RA. This is reflected in a failure to downregulate stem cell genes or upregulation of differentiation specific genes.

Several studies in both yeast and mammals have demonstrated that H3K36me3 is a marker for recent transcribed chromatin (Bernstein et al., 2005; Morris et al., 2005; Muller et al., 2007; Pokholok et al., 2005; Vakoc et al., 2006). Most of the promoters contain H3K4me3 and RNA polymerase II, but only a subset of genes with H3K36me3 transcribe (Guenther et al., 2007), suggesting H3K36me3 is required for efficient transcription. I observed an increase of H3K36me3 during differentiation. Thus, knockdown of HypB may impair the proper transcription of genes that normally need to be activated during differentiation. Alternatively, the observed change in gene expression is just consequence of differentiation defects.

3.3 HypB facilitates ES cell differentiation: a role at the promoter and enhancer of *Oct4*?

HypB knockdown ES cells show a similar phenotype as *Mbd3*-null ES cells. *Mbd3*-null ES cells also failed to repress stem cell markers and activate differentiation markers, which are analyzed in HypB knockdown ES cells. Mbd3 is found to be recruited to the promoter of *Oct4* and required for the repression of *Oct4* during differentiation. Mbd3 is a component of the NuRD complex, which contains another protein CHD3. CHD3 has been identified as an H3K36me3 binding protein. In addition, the proximal enhancer (PE) of *Oct4* is found to be required for downregulation of *Oct4* expression after differentiation induced by removal of LIF. *Oct4/lacZ* transgene expression was not down regulated when the PE is absent, but was down regulated when the PE is present, after the transgenic ES clones were induced to form embryo bodies (Ovitt and Scholer, 1998). The transgenic ES clone with the PE deletion construct showed similar phenotype of *Oct4* expression as *Mbd3*-null and HypB knockdown cells. Thus, HypB may have a direct effect on the promoter and enhancer for the downregulation of *Oct4* expression.

In yeast, there is some evidence to support a role of H3K36 methylation on the promoter of genes. It was reported that SET2 is a repressor of *GAL4* transcription (Landry et al., 2003). The *GAL4* promoter, which does not contain a TATA box, is regulated by three *cis*-acting elements: a UAS necessary for maximal *GAL4* expression, an upstream sequence that is essential for *GAL4* expression, and Mig1 binding sites that mediate glucose repression of *GAL4* expression. UAS deleted mutation have a low

basal level of *GAL4* expression. SET2 functions in this basal repression of *GAL4* as H3K36 methylation is present on the promoter. Moreover, SET2 mutant significantly increases expression of a UAS-less *GAL4* gene. In yeast, SET2 has an opposite role to yFACT in TBP binding on the promoter of genes (Biswas et al., 2006). TBP, although named TATA-binding protein, is also involved in regulating TATA less promoter (Ohbayashi et al., 2003; Pugh and Tjian, 1991; Weis and Reinberg, 1997).

Like the yeast *GAL4* gene, *Oct4* has a TATA-less promoter, and regulated by three *cis*-acting elements, the distal enhancer, the proximal enhancer, and proximal promoter. It is known that the distal enhancer is required for *Oct4* expression and the proximal enhancer is required for repression of *Oct4* (Ovitt and Scholer, 1998). *Oct4*ΔDE/*lac*Z reporter was repressed in ES cells (Yeom et al., 1996). To investigate whether HypB regulates *Oct4* gene expression in mammalian system by a similar regulatory way as yeast SET2 regulates *GAL4*, it will be informative to perform HypB knockdown in *Oct4*ΔDE/*lac*Z reporter transfected ES cells to test if the reporter can be reactivated upon reduction of HypB.

Chromatin immunoprecipitation on *Oct4* locus showed that H3K36me3 is highly enriched in the coding region in undifferentiated ES cells. H3K36me3 clearly decreases on *Oct4* coding region after its repression upon differentiation. In contrast, a small increase of H3K36me3 on the promoter of *Oct4* was observed after differentiation. The levels, however, resemble the amount of H3K36me3 at 3' end of gene in the repressed condition.

As discussed above, Mbd3 is recruited to the promoter of *Oct4*, but it only transiently present around 36 hours after differentiation induced by RA. It is tempting to suggest a model (Figure 56) in which H3K36 facilitates or reinforces the binding of Mbd3 through CHD3/NuRD for repressing of *Oct4* during differentiation, but this hypothesis needs further verification. The preliminary data did not reveal an increase of H3K36me3 at this time point. Ultimately, it is needed to show that the recruitment of Mbd3 to the *Oct4* promoter during differentiation is reduced upon HypB knockdown. So far, attempts to reproduce the data of recruitment of Mbd3 to the promoter of *Oct4* (Gu et al., 2006) failed. It is still possible that Mbd3/NuRD function in the coding region as proposed above, which results in the similar phenotypes between *Mbd3*-null and HypB

knockdown ES cells. Alternatively, HypB may function in a distinct way, independent of its activity towards histone.

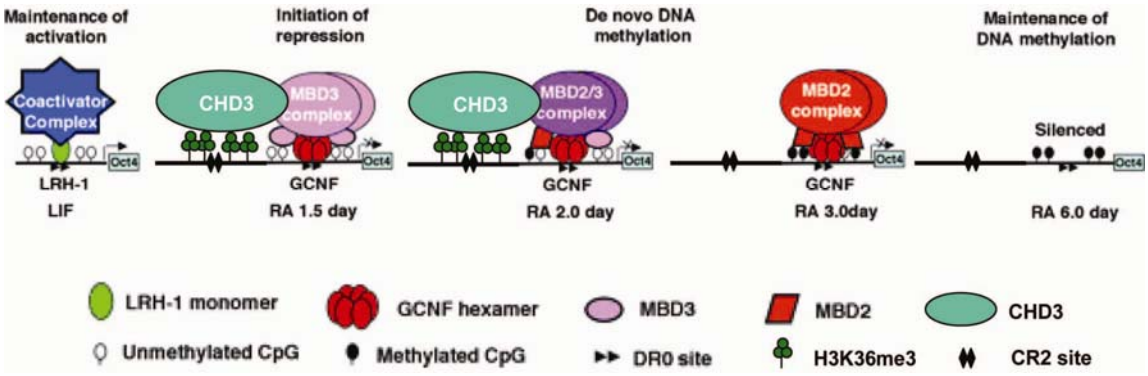

Figure 56. Working model of H3K36me3 on the promoter of Oct4 during differentiation

HypB-mediated H3K36me3 is involved in facilitating or reinforcing the recruitment of the NuRD complex to the *Oct4* promoter after differentiation. CR2: conserved region 2. (Adapted from Gu et al., 2006)

To understand the molecular mechanism of HypB and H3K36me3 in differentiation, genome wide mapping of H3K36me3 and HypB during differentiation by ChIP-chip or ChIP-seq, transcription profiling gene expression during differentiation in HypB knockdown cells, and identification of HypB interaction partners in undifferentiated versus different stages of differentiation would be worthwhile to be performed.

3.4 The dynamics of H3K36 methylation on chromatin

The immunofluorescence staining showed that H3K36me1 and H3K36me2 not only localized in euchromatin, but also showed heterochromatic enrichment at DAPI dense region (Figure 15). In undifferentiated ES cells, these two H3K36 methyl markers were present in euchromatin at low levels. Upon differentiation, global levels increased and correlated inversely with the presence of the stem cell marker Oct4 (Figure 22). Importantly, H3K36me1 and H3K36me2 gained heterochromatic enrichment upon differentiation (Figure 23). Accordingly, in mouse oocytes and early embryos, no heterochromatic localization of H3K36me1 or H3K36me2 was observed. The staining of H3K36me2 in oocyte and early embryo is shown in Figure 57. This may indicate that these two methyl markers acquire heterochromatic enrichment during differentiation at specific stage of early embryonic development and play an important role during early development.

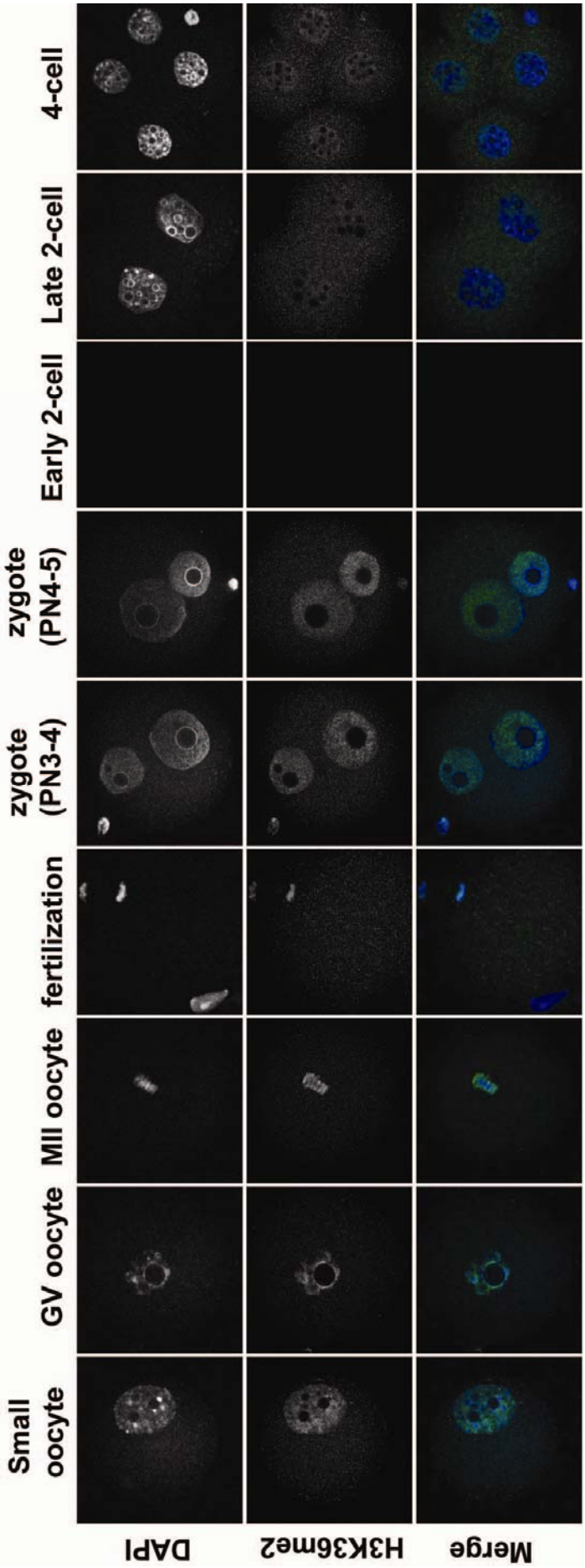

Figure 57. Localization of H3K36m2 in mouse oocyte and early embryo

H3K36me2 is maternally transmitted. The paternal genome acquires H3K36me2 prior to replication in early zygotes.

No clear staining at pericentric heterochromatin is observed (DAPI-enriched regions surounding DAPI-pale nucleoli).

Courtesy of Mareike Puschendorf

The putative candidates of H3K36 methyltransferase Nsd2 and Nsd3 showed a weak *in vitro* HMT activity towards H3. Interestingly, when *Nsd2* and *Nsd3* were overexpressed using *EGFP* fusion constructs in NIH3T3 cells, both proteins showed localization at heterochromatin (Figure 58). The presence of PWWP domains in Nsd2 and Nsd3 may explain their heterochromatic localizations since the similar PWWP domain in Dnmt3a and Dnmt3b targets them to heterochromatin (Chen et al., 2004). The localization of Nsd2 and Nsd3 protein is currently under investigation with the antibodies generated in our lab.

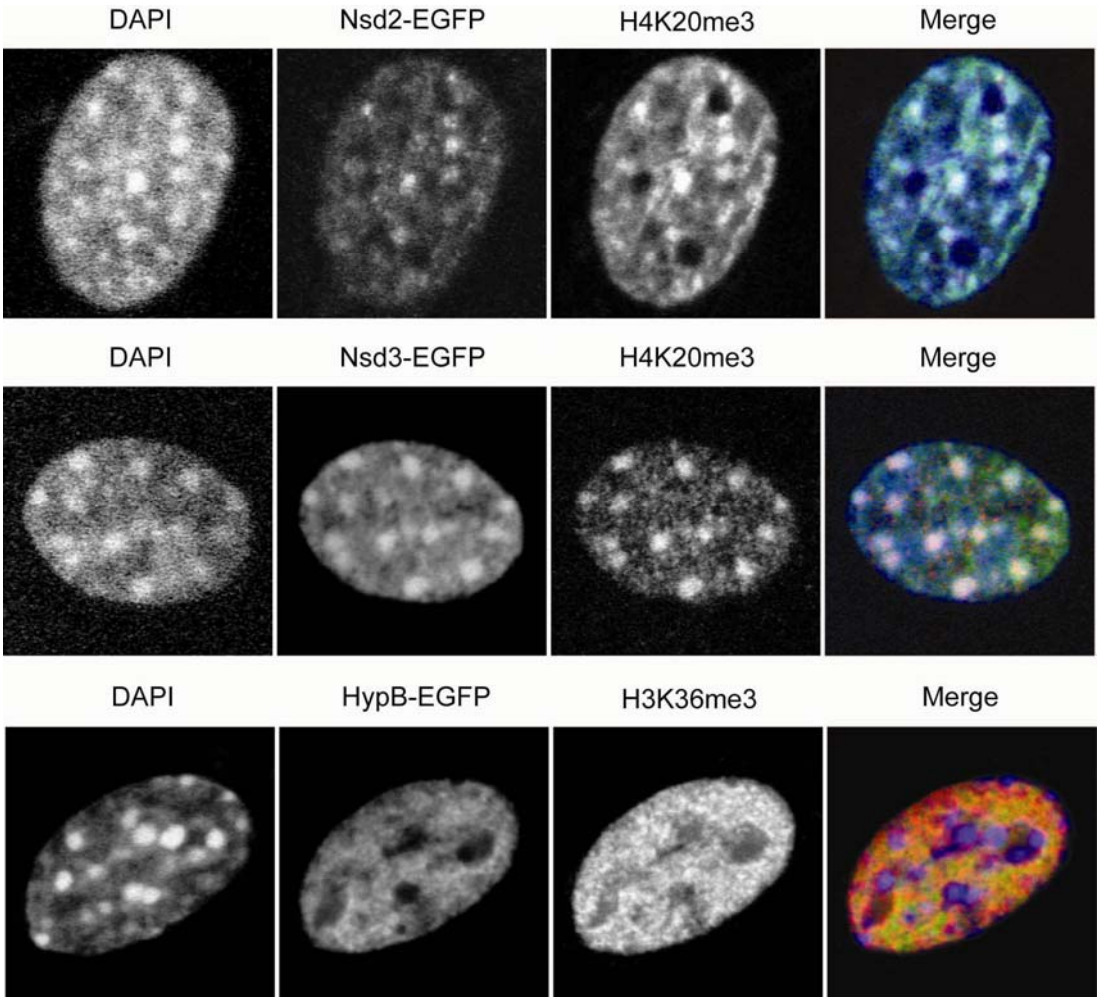

Figure 58. Localization of overexpressed Nsd2, Nsd3, and HypB

Nsd2-EGFP and Nsd3-EGFP localized to euchromatin and Heterochromatin. HypB-EGFP only localized to euchromatin.

It still needs to be clarified whether Nsd2 and Nsd3 are responsible for H3K36me1 and/or H3K36me2 at heterochromatin in mammals and if the gain of H3K36me1 and

H3K36me2 at heterochromatic region has functional significance during embryo development. A PWWP domain is also present in the Nsd1 protein; it might recruit Nsd1 to heterochromatin as well.

The question remains why H3K36me1 and H3K36me2 are not present at pericentric heterochromatin in undifferentiated ES cells. One possibility is that the chromatin state or associated factors in undifferentiated ES cells block the recruitment of the Nsd proteins. The second possibility is that the pericentric heterochromatin undergoes some kind of maturation during differentiation that allows the targeting of the Nsd proteins in differentiated cells.

In contrast, H3K36me3 only localizes in euchromatin. In undifferentiated ES cells, H3K36me3 is present in euchromatin at low levels. Upon differentiation, global levels increased and correlated inversely with the presence of the stem cell marker Oct4. Overexpressed HypB-EGFP in NIH3T3 cells localized to euchromatin and correlate with the localization of H3K36me3 (Figure 58). We can not completely exclude the activity of HypB for H3K36me1 and H3K36me2 *in vivo*, although only decrease of H3K36me3 was observed in HypB knockdown cells.

Interestingly, H3K36me3 labels oocyte and maternal genome in early zygotes (Figure 59), reflecting establishment during transcription occurring in growing oocyte and stable inheritance of this marker to the embryo. H3K36me3 is lost during first round of replication, transition form early to late zygote. H3K36me3 is newly established in late 2-cell embryo, concurrent with genome wide activation of transcription. It will be interesting to study the function of H3K36me3 during early embryo development by RNAi or gene targeting.

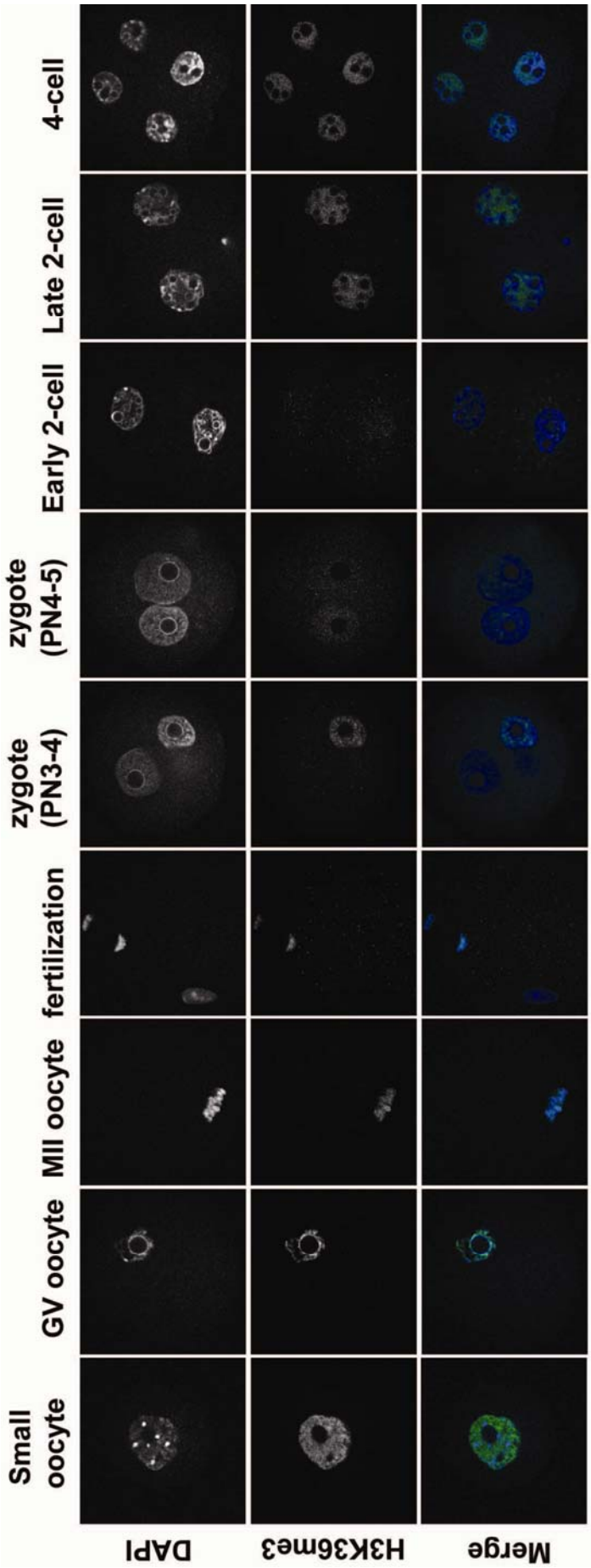

Figure 59. Localization of H3K36m3 in mouse oocyte and early embryo

H3K36me3 is maternally transmitted. The signal intensity of H3K36me3 is reduced at time of replication in one cell embryo and remains low in early 2-cell embryo. The H3K36me3 levels strongly increase at late 2-cell embryo coinciding with the time of zygotic genome activation. (Courtesy of Mareike Puschendorf)

3.5 Additional roles of H3K36 methylation

H3K36 methylation is implicated in other processes. In yeast, the SET2 deletion strain showed a mild phenotype to irradiation, implicating the role in DNA repair (Game et al., 2006). The increase of antisense transcripts in the SET2 deletion strains is generally associated with hypersensitivity to DNA-damaging agents (Nicolas et al., 2007). MRG15 is present in the Tip60 complex, which selectively acetylates histone variant and results in histone variant exchange at DNA lesions (Kusch et al., 2004). Immunostaining in irradiated U2OS cells show that H3K36me3 is clearly excluded from phosphorylated H2A.X foci, a marker of double stand break sites (Figure 60). Further studies need to address if H3K36 methylation is indeed involved in DNA repair.

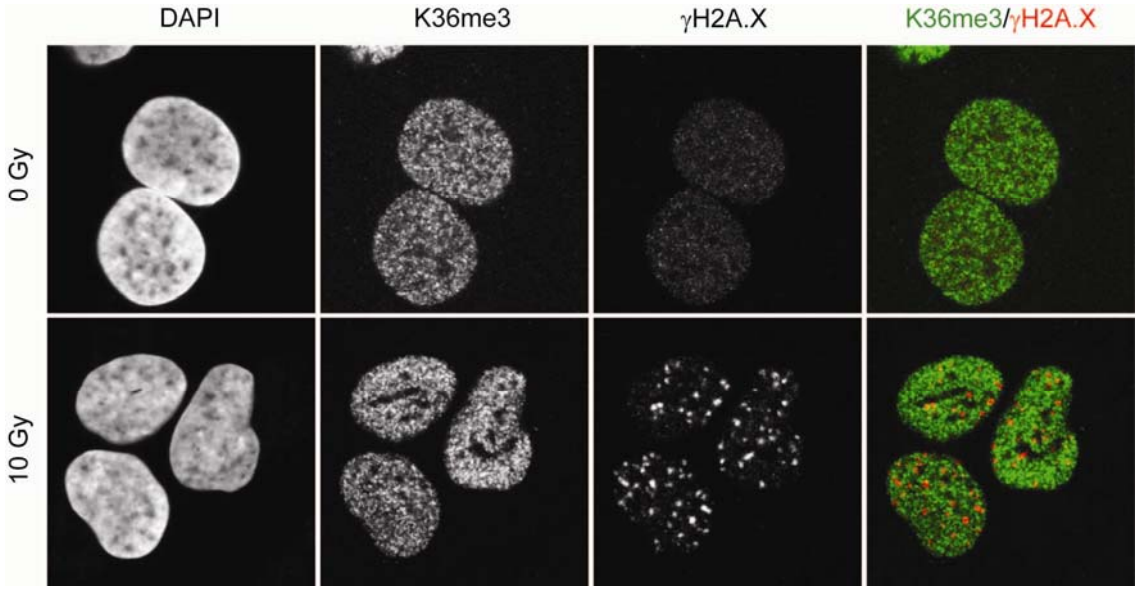

Figure 60. H3K36me3 was excluded from DNA double stands break

U2OS cells were exposed to 10 Gy ionizing irradiation and examined 15 min later by immunofluorescence.

In *C. elegans*, H3K36me2 methylated by Mes-4 participates in X-linked gene silencing, together with Mes-2/-3/-6. H3K36 methylation may function in *Drosophila* dosage compensation because of its crosstalk with H4K16 acetylation (Bell et al., manuscript in preparation) and targeting of MSL3 was found toward the 3' end of the gene and dependent on gene transcription (Kind and Akhtar, 2007). The distribution patterns of H3K36 methylation on autosomes and sex chromosomes in mammalian cells are unknown yet. It will be very interesting to investigate if the H3K36 methylation in

mammals functions in X-linked gene silencing by a similar mechanism in *C. elegans*, and in dosage compensation by a similar mechanism in *Drosophila*.

3.6 H3K36 methyltransferase and cancers

All three genes *NSD1*, *NSD2*, and *NSD3* have been implicated in cancer. Rearrangements of the genetic sequences of *NSD1* and *NSD3* that result in truncations are associated with acute myeloid leukemia, and a fusion of *NSD2* that is thought to cause overexpression of the gene is associated with multiple myeloma (Schneider et al., 2002). *NSD3* is also rearranged in several cancerous cell lines and primary breast carcinomas (Schneider et al., 2002). Recently, it has been reported that NUP98-NSD1 links H3K36 methylation leukaemogenesis. It binds to genomic elements adjacent to *HoxA7* and *HoxA9*, maintains H3K36 methylation and prevents EZH2-mediated H3K27 methylation for transcription repression of *Hox-A* locus during differentiation (Wang et al., 2007).

HYPB has been identified as a candidate gene deleted at 3p21.31 in multiple myeloma by array comparative genomic hybridization (Carrasco et al., 2006). HYPB is downregulated in leukemia (Figure 61). It will be highly interesting to investigate if HYPB may act as a tumor suppressor gene and if downregulation of HYPB promotes cancer stem cell self-renewal.

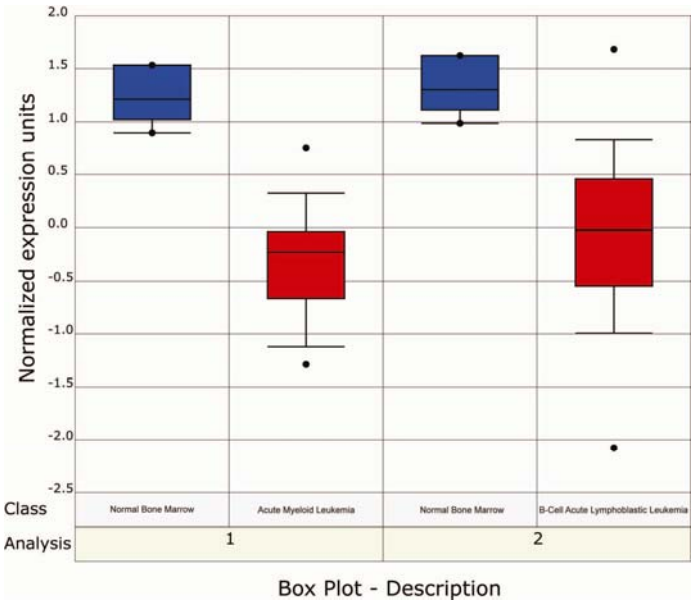

Figure 61. Downregulation of HYPB in leukemia

Data was obtained from http://www.oncomine.org/main/mainx.jsp

3.7 Conclusion

HypB is identified as a major H3K36 trimethylase *in vivo*. H3K36me3 localizes to euchromatin. The levels of H3K36me3 are low in undifferentiated ES cells and increase in differentiating ES cells identified by the low levels of Oct4. HypB knockdown cells fail to repress the stem cell-specific genes and activate differentiation-specific gene after differentiation. Thus, HypB-mediated H3K36me3 has an important role in facilitating ES cell differentiation.

H3K36me1 and H3K36me2 localize to both euchromatin and heterochromatin in differentiating ES cells and terminally differentiated NIH3T3 cells. In contrast, these two methyl markers only localize to euchromatin in undifferentiated ES cells and in early mouse embryo, suggesting heterochromatic gain H3K36me1 and H3K36me2 during differentiation.

The mechanism of H3K36me3 in facilitating ES cell differentiation need to be further dissected. Experiments will be performed to investigate whether the known and the potential pathways discussed above are involved in the observed phenotype of HypB knockdown cells. Genome wide mapping of H3K36me3 and HypB during differentiation by ChIP-chip or ChIP-seq, transcription profiling gene expression during differentiation in HypB knockdown cells, and identification of HypB interaction partners in undifferentiated versus different stages of differentiation will be informative to understand the function of HypB and H3K36me3.

CHAPTER 4. MATERIALS AND METHODS

4.1 Materials

4.1.1 Cell lines and culture media

NIH3T3 mouse fibroblasts, ATCC CRL-1658. Cells are cultured in DMEM supplemented with 10% FCS.

293 human embryonic kidney carcinoma, ATCC CRL 1573. Cultured in DMEM supplemented with 10% FCS.

U2OS human bone osteosarcoma epithelial cells, obtained from the cell collection of the Antoine Peters laboratory. Cells are cultured in DMEM supplemented with 10% FCS.

CCE mouse embryonic stem (ES) cell line, obtained from the cell collection of the Antoine Peters laboratory.

PGK12 mouse embryonic stem (ES) cell line, obtained from the cell collection of the Antoine Peters laboratory.

LF2 mouse embryonic stem (ES) cell line, obtained from the cell collection of the Antoine Peters laboratory.

Ecophoenix helper-free retrovirus producer lines, obtained from the cell collection of the Antoine Peters laboratory.

Dulbecco's Modified Eagles Medium (DMEM)	Invitrogen/Gibco BRL
Fetal Calf Serum (FCS)	Invitrogen/Gibco BRL
Fetal Calf Serum (FCS) (for ES maintenance)	Invitrogen/Gibco BRL
Penicillin G/Streptomycin (100x)	Invitrogen/Gibco BRL
Glutamine (200 mM)	Invitrogen/Gibco BRL
Trypsin/EDTA (1x)	Invitrogen/Gibco BRL
LIF	Peters Lab
Sodium Pyruvate	Invitrogen/Gibco BRL
Non-Essential Amino Acids	Invitrogen/Gibco BRL
β-Mercaptoethanol	Invitrogen/Gibco BRL

4.1.2 Antibodies

Goat α -mouse HRP	Sigma
Sheep α -rabbit HRP	GE healthcare
Rabbit α -goat HRP	Santa Cruz
Goat α -rabbit-Alexa 488 or 565 conjugates	Molecular Probes
Goat α -mouse-Alexa 488 or 565 conjugates	Molecular Probes
Anti-H3K36me1	Abcam
Anti-H3K36me2	Upstate
Anti-H3K36me3	Abcam
Anti-H3K4me3	Upstate
PCNA	Dako
H3 serine 10 phosphorylation	Upstate
Oct4	BD pharmingen, Santa cruz
Mbd3	Santa cruz
Rabit IgG	Sigma

4.1.3 Histone peptides

H3 (1-20)	ARTKQTARKSTGGKAPRKQL-cys
H3 (1-20) K4L	ARTLQTARKSTGGKAPRKQL-cys
H3 (1-20) K9L	ARTKQTARLSTGGKAPRKQL-cys
H3 (19-38)	QLATKAARKSAPATGGVKKP-cys
H3 (19-38) K27L	QLATKAARLSAPATGGVKKP-cys
H3 (19-38) K36L	ARKSAPATGGVLKPHRYRPGT-cys
H3 (19-38) K37L	ARKSAPATGGVKLPHRYRPGT-cys
H3 (19-38) K36LK37L	ARKSAPATGGVLLPHRYRPGT-cys
H3 (19-38) K36me1	ARKSAPATGGVK (Me) KPHRYRPGT-cys
H3 (19-38) K36me2	ARKSAPATGGVK (Me2) KPHRYRPGT-cys
H3 (19-38) K36me3	ARKSAPATGGVK (Me3) KPHRYRPGT-cys
H4 (12-31)	KGGAKRHRKVLRDNIQGITK-cys
H4 (12-31) K20L	KGGAKRHRLVLRDNIQGITK-cys
H1a (18-33)	PAAAKKTKKPAKAAAP-cys
H1b (18-33)	SPAKKKTTKKAGAAKR-cys
H1c (18-33)	APAKKKAAKKPAGVRR-cys

H1d (18-33)	TPVKKKAKKTGAAAGK-cys
H1e (18-33)	TPVKKKARKAAGGAKR-cys
H1t (20-35)	PSSKRRGKKPGLAPAR-cys
H10 (12-27)	KPKRAKASKKSTDHPK-cys

4.1.4 RT-PCR primers of gene expression

Human genes:

Gene	Primer	Sequence (5' to 3')
NSD1	F	TCCAGAAGTACCCACCCACTG
	R	GCGCATCAACCGACCTTTAG
NSD2	F	TGAAGCTGAGGACACACCCA
	R	GTTGCTGCCTGGCTCTTGAG
NSD3	F	CGGGTACGAGAGTATAAAGG
	R	CTGAGGTCGGGGTTTCCGAA
HypB	F	AAAATGGACTGTGAACGGACAA
	R	TGGCTGATCCGCAGAAACA
GAPDH	F	TTGCCATCAATGACCCCTTCA
	R	CGCCCCACTTGATTTTGGA

Mouse genes:

Gene	Primer	Sequence (5' to 3')
Nsd2	F	GTGCACGCCAGTATCATGTAC
	R	CTCTCAATCTCCCCGAAATAG
HypB	F	GAGGTGTATCGGATTCCAAAG
	R	TGACAGAGAGCCCCGTCGTTT
Oct4	F	CACACTCTACTCAGTCCCTTTTCCT
	R	CAAAGCTCCAGGTTCTCTTGTCTAC
Rex1	F	GTCCTTTTGATGGCTGCGAG
	R	TCCCCAGTGCCTCTGTCATT
Nanog	F	CCTCCAGCAGATGCAAGAACTC
	R	CTTCAACCACTGGTTTTTCTGCC
Sox2	F	CGAGATAAACATGGCAATCAAATG

	R	AACGTTTGCCTTAAACAAGACCAC
Brachyury	F	GCTCTAAGGAACCACCGGTCATC
	R	ATGGGACTGCAGCATGGACAG
Gata6	F	GGGCCTTGTCTGCTAAGGAAG
	R	CCACGAACGCTTGTGAAATG
Flk1	F	CTGTGGCGTTTCCTACTCCT
	R	AGGAGCAAGCTGCATCATTT
Fgf5	F	CTGTATGGACCCACAGGGAGTAAC
	R	ATTAAGCTCCTGGGTCGCAAG
Nestin	F	CAGCAACTGGCACACCTCAA
	R	CCCAAGGAAATGCAGCTTCA
Dppa3	F	AGGCTCGAAGGAAATGAGTTTG
	R	TCCTAATTCTTCCCGATTTTCG
Gapdh	F	ATGTTCCAGTATGACTCCACTCACG
	R	GAAGACACCAGTAGACTCCACGACA

4.1.5 RT-PCR primers on *Oct4* locus of ChIP

Region	Primer	Sequence (5’ to 3’)
R1	F	CATTATAGGTGTGGCATTCCGCATCTG
	R	TGCCACAAACCACCTGTATTTTAGAACCA
DE	F	GGAACTGGGTGTGGGGAGGTTGTA
	R	AGAAGATTAAGGAAGGGCTAGGACGAGAG
R2	F	TGCTCTGGGCTTTTTGAGGCTGTGTGATT
	R	TGGCGGAAAGACACTAAGGAGACGGGATT
PP	F	TGGGCTGAAATACTGGGTTC
	R	GTCCTTACAGCCCACTCAG
R3	F	CCCTGCAGAAGGAGCTAGAACA
	R	CCTACATTAAGAGCCGTGAGATCAG
R4	F	CACACTCTACTCAGTCCCTTTTCCT
	R	CAAAGCTCCAGGTTCTCTTGTCTAC

4.1.6 siRNA and shRNA templates

shRNA/pLKO.1 for human genes (Sigma):

NSD1: TRCN0000061353 CCGAACACTTTCCCACTGTTA

NSD2: TRCN0000019816 CCTCTCTTTGAATCTTCCATT

NSD3: TRCN0000015613 CGAGAGTATAAAGGTCATAAA

TRCN0000015615 CGAGAATATCATGTCCAGTTT

HYPB: TRCN0000003030 CCTGAAGAATGATGAGATAAT

TRCN0000003031 GCAAGTAAAGCCTGTCCTCAA

TRCN0000003032 GCCCTATGACTCTCTTGGTTA

TRCN0000003033 CTTACCTGTCTGGAACTCATA

Stealth RNA for human *HYPB* (Invitrogen):

Sense: AUAAGGAGCAGGAGAACACACUGGG

Antisense: CCCAGUGUGUUCUCCUGCUCCUUAU

shRNA/MSCV-LMP for mouse *HypB*:

CGGAAATAAGCAGCAATTCAA

Stealth RNA for mouse *HypB* (Invitrogen):

Sense Sequence: GCACCGGAAAUAAGCAGCAAUUCAA

Antisense Sequence: UUGAAUUGCUGCUUAUUUCCGGUGC

4.1.7 Plasmids

Map of the general plasmids, pGEX-6P-1, pMSCV-LMP, and pCX-EGFP are shown in Appendix.

shRNA expression plasmids:

pLKO.1-scramble
pLKO.1-shNSD1
pLKO.1-shNSD2
pLKO.1-shNSD3
pLKO.1-shHypB

pMSCV-LMP-shHypB

Eukaryotic expression plasmids:

pCX-Nsd2-EGFP

pCX-Nsd3-EGFP

pCX-HypB-EGFP

4.1.8 Enzymes

Enzymes were obtained form New England Biolabs (NEB), Roche Diagnostics, or Promega and were used in the buffer combinations recommended by the manufacturer.

Restriction enzymes	NEB, or Roche
Calf intestine phosphatase (CIP)	NEB
T4 DNA ligase	NEB
Pfu DNA polymerase	Promega
Pfx DNA polymerase	Invitrogen
Taq DNA polymerase	Promega
Tag plus DNA polymerase	Statagene
RNaseA	Roche
RNase Inhibitor RNasin®	Promega

4.1.9 Bacteria and media formulations

Top10:	Invitrogen
Top10F':	Invitrogen
DH5α:	BD/Takara.
BL21(DE3):	Stratagene
BL21 (DE3) pLysS:	Invitrogen

Bacteria are cultured in Luria Bertani (LB) medium at 37°C in presence of the corresponding selection antibiotics.

4.1.10 Kits

Bradford protein determination assay	BioRad
ECL Western Blotting Kit	GE Healthcare
LipofectAMINE 2000 transfection reagent	Qiagen
pcDNA/HisMax TOPO TA expression Kit	Invitrogen
THERMOSCRIPT RT-PCR System	Invitrogen
BrdU Flow Kits	BD Biosciences

Plasmid Mini, and Maxi Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
RNeasy Mini Kit	Qiagen
Ni-NTA Superflow	Qiagen
Talon metal affinity resin	Takara

4.2 Methods

4.2.1 Cell culture:

293, U2OS, Ecophoenix, and NIH3T3 cells were cultured in DMEM supplemented with 10% FCS. Mouse ES cells CCE, PGK12, and LF2 were cultured in High glucose DMEM supplemented with 15% FCS (for ES maintenance), 10 ng/mL LIF (Leukemia Inhibitory Factor), 1 mM Sodium Pyruvate, 0.1 mM Non-Essential Amino Acids, 1mM β-Mercaptoethanol. ES cells were cultured on gelatinized tissue culture dishes. ES cells were passaged every 2-3 days at a ratio of 1/20 - 1/40 by washing with PBS, dissociating with 0.1 trypsin-EDTA for 5 min at 37°C, and resuspending in ES medium. Media were change daily. All cell culture media are supplemented with 100U/ml penicillin G, 100μg/ml streptomycin, and 2mM glutamine and all cells are cultured under 5% CO2 in atmospheric air at 37°C unless otherwise stated.

4.2.2 shRNA plasmid transfection

Transfection was performed according to manufacture's protocol (Invitrogen's Lipofectamine 2000). Briefly, cells (2 x 10^5) were seeded in 6-well plates 1 day before transfection. 2μg shRNA plasmids were transfected by 5 μl Lipofectamine 2000 in 2 ml medium per well.

4.2.3 Stealth RNA transfection

Transfection was performed according to manufacture's protocol (Invitrogen's Lipofectamine 2000). Briefly, cells (2 x 10^5) were seeded in 6-well plates 1 day before transfection. Transiently transfected cell were either harvested at day 3 without

selection or at day 5 with puromycin selection for analysis. 20nM stealth RNA was transfected by 5 μl Lipofectamine 2000 in 2 ml medium per well.

4.2.4 siRNA and shRNA vector design

shRNA in pLKO.1 vector targeting human *NSD1*, *NSD2*, *NSD3*, and *HYPB* genes were purchased from Sigma. Stealth RNA (25 base-pair blunt siRNA) targeting human and mouse *HypB* gene were purchased from Invitrogen. A miR30-based shRNA template targeting mouse HypB was cloned into MSCV-LTRmiR30-PIG (LMP) vector according to the manufacture's protocol (Open Biosystems).

4.2.5 Retroviral infection of shRNA

To produce viruses, 5 x 10^6 ecophenix cells were seeded on 10-cm dishes 12-24 hours before transfection. 24μg shRNA plasmids were transfected by 60 μl Lipofectamine 2000 per well. After 6 hours incubation at 37ºC, the medium was change to ES medium. 24 hours after transfection, cells were transferred to 32ºC and cultured for 48 hours. The supernatant was collected and filtered by 0.45-μm filter. The supernatant was supplemented with 5μg/ml polybrene and were ready to use or stored at -80ºC.

To infect ES cells, cells were plated in 6-well dishes 12 hours before infection. The medium was replaced by 2.5 ml supernatant containing viruses and spin at 1000g for 1 hour. The cells were incubated for 6 hours at 37ºC. Then the supernatant was replaced by fresh ES medium and cultured for shRNA expression. Infected cells harvested at 72 hours after infection or after selection by puromycin containing medium were used for gene expression analysis. To establish a stable pool of overexpressing shRNA, ES cells were sequentially infected, ~12 hours apart. GFP positive cells sorted by FACS were put in culture under puromycin selection.

4.2.6 Staining of live cells with Hoechst 33342

Hoechst 33342 was added to 1 x 10^6 cells/ml cell suspension in FACS buffer (PBS without Ca^{2+} and Mg^{2+} with 3% FCS) to obtain a final fluorophore concentration of 2.0 μg/ml. Cells were incubated 20 min at 37ºC. Cells were sorted by flow cytometer for UV excitation at 340-380 nM.

4.2.7 Cell proliferation and cycle analysis

To assay proliferation, CCE cells were seed at 2.8 x 10^4 cells (in triplicate) in gelatinized 6-well dishes and transfected with stealth RNA at the second day (day 0). Cells were counted by CASY cell counter from day 0 to day 3.

The BrdU incorporation study was performed according to manufacturer's protocol (BD, BrdU Flow Kit). Briefly, CCE cells transfected by stealth RNA were incubated with BrdU for 1 hour at 37ºC. The incorporated BrdU was stained with specific anti-BrdU fluorescent antibodies. 7-amino-actinomycin D (7-AAD) was used as a staining dye for total DNA to define G1, S, or G2/M phases of cell cycle. The cells were analysed by flow cytometry.

4.2.8 Immunoflurorescense staining

ES cells were trypsinized and placed on poly-L-lysine coated coverslips for 10 min to attach. Cells were fixed with 2% paraformaldehyde in PBS (pH 7.4), permeabilized in 0.1% Triton-X100 in 0.1% sodium citrate and blocked for 30 min in 0.1% Tween-20 in PBS containing 2.5% BSA and 10% normal goat serum at RT. Fixed cells were blocked 30 minutes at RT in 0.1% Tween-20 in PBS containing 2.5% BSA and 10% normal goat serum, and were then incubated with primary antibodies in blocking solution for 2 h at 4ºC. Double antibody staining was accomplished by mixing appropriate different primary and different secondary antibodies for simultaneous incubation. Cells were washed three times for 10 min in 0.1% Tween-20 in PBS containing 0.25% BSA before application of secondary antibodies. For detection, anti-rabbit IgG-Alexa 488, anti-mouse IgG-Alexa 568 (Molecular Probes) secondary antibodies were diluted 1:500 in blocking solution and cells were incubated for 1 h at RT followed by three washing steps a 10 min in 0.1% Tween-20 in PBS in the dark. Cells were mounted in Vectashield containing DAPI (Vector laboratories).

293 cells were grown on poly-D-lysine treated coverslips, U2OS and NIH3T3 cells were grown on normal coverslips 1 day before staining. The procedures were performed identical as described for ES cells.

4.2.9 GST fusion protein cloning, expression, and purification

DNA fragments coving catalytic domain (pre-SET, SET, and post-SET domain) of *Nsd1*, *Nsd2*, *Nsd3*, and *HypB* were cloned into pGEX-6P-1 vector. Transformation of plasmid

into BL21 bacterial was performed according to the manufacture's protocol (Stratagene). Overnight bacterial culture from single colony was diluted 100x and cultured at 37ºC for ~2 hours to OD 0.6-0.8, then induced by 0.1 mM IPTG for 6 hours at 28ºC. Bacterial pellets were suspended in lysis buffer (1% Triton X-100, 50mM Tris-HCl pH7.5, 150nM NaCl, 1mM DTT, protease inhibitor cocktail) with 0.25mg/ml lysozyme on ice for 20 min and then sonicated 3x 10 seconds with 50%. Recombinant proteins were purified using Glutathione Sepharose 4 Fast Flow according to the manufacture's protocol (Amersham Pharmacia Biotech).

4.2.10 Histone methyltransferase assay

In a final reaction volume of 20 μl, the protein sample was incubated at 30ºC for 1 hr in a reaction buffer containing 50 mM Tris-HCl, 4 mM DTT, and 1 μM ^{3}H-labeled *S*-adenosyl-L-methionine [^{3}H]SAM (GE Healthcare). Different pH of Tris-HCl and different Mg^{2+} concentration was titrated. 2μl of 10mM histone tail peptides, 2μg of native histone mixture (Roche), histone octamers, or oligonucleosomes was used as substrates. The reaction was stopped by addition of SDS sample buffer, and the reaction mixtures were fractionated by 12% or 15% SDS-PAGE. Separated peptides or histones were then transferred to a PVDF membrane and visualized by Ponceau staining. The membrane was exposed to phosphoimager screen overnight or longer and scanned by Typhoon 9200.

4.2.11 RNA isolation and reverse transcription

Total RNA was isolated by Trizol according to the manufacturer's instructions (Invitrogen). 1.25μg total RNA was reverse transcribed with oligo(dT) primers for 60min at 50ºC according to the manufacturer's protocol (THERMOSCRIPT RT-PCR System, Invitrogen).

4.2.12 *In vitro* ES cell differentiation

Differentiation media consisted of either ES medium without LIF or ES medium without LIF and supplemented with 0.1 μM all-trans retinoic acid. For RNAi experiment, medium was changed every day. For time course of differentiation, ES cells were plated on 10-cm gelatinized tissue culture dishes. In differentiated medium with LIF, ES cells were

seeded at a ratio of 1/2 for 36 hours differentiation, 1/3 for 48 hours differentiation, 1/10 for 72 hours differentiation, 1/40 for 96 hours differentiation, and 1/120 for 144 hours differentiation; In ES medium without LIF and supplemented with 0.1 μM all-trans retinoic acid, 1/2 for 36 hours differentiation, 1/3 for 48 hours differentiation, 1/5 for 72 hours differentiation, 1/10 for 96 hours differentiation, and 1/40 for 144 hours differentiation. Media were change every 2 days. Control cells were harvested 2 days after plating in +LIF media. Cells were harvested for total RNA isolation, cell extraction, immunostaining or chromatin immunoprecipitation.

4.2.13 Chromatin immunoprecipitation (ChIP)

Cells were crosslinked by adding 11x fixation solution (50mM Hepes at pH7.5, 1mM EDTA, 0.5mM EGTA, 100mM NaCl, 11% Formaldehyde) for 10 min at RT. Crosslinking was stopped by adding 11x glycine (1.375M Glycine) to a final concentration of 125mM. Cells were lysed with SDS buffer (50mM Tris-HCl at pH8.0, 1% SDS, 10mM EDTA) to 20 x 10^6 cells/ml in the presence of 1x protease inhibitor (complete, Roche) for 5 min at RT and diluted to 5 x 10^6 cells/ml with dilution buffer (20mM Tris-HCl at pH8.0, 150mM NaCl, 1% triton X-100, 2mM EDTA). Cell lysates were sonicated to a genomic fragment length of 500-1000 base pair. Cell lysates were diluted 2.5x to 0.1% SDS and precipitated by primary antibody with rotation overnight at 4ºC. Overnight pre-blocked protein A beads were added to cell lysates and rotated at 4ºC for at least 4 hours. Beads were washed 3x with low salt washing buffer (20mM Tris-HCl at pH8.0, 150mM NaCl, 1% triton X-100, 0.1% SDS, 2mM EDTA), 1x high salt washing buffer (20mM Tris-HCl at pH8.0, 500mM NaCl, 1% triton X-100, 0.1% SDS, 2mM EDTA),and TE (10mM Tris-HCl at pH8.0, 1mM EDTA).

4.2.14 Real time PCR

Real time PCR was performed with ABI prism 7000 Sequence Detection System (Applied Biosystems). For gene expression analysis, real time PCR was performed using qPCR MasterMix Plus for SYBR green I QGS (Eurogentec) or SYBR green PCR master mix (Applied Biosystems). 5μl 20x diluted cDNA and 5μl genomic DNA was used for 50 μl reaction. Differences for expression level and ChIPed DNA were calculated according to the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

4.2.15 Histone isolation

1~2 x 10^6 cells were suspended in 1 ml lysis buffer (10mM Tris-HCl at pH6.5, 1% Triton X-100, 20mM $MgCl_2$, 50mM sodium bisulfite, 8.6% Sucrose) and homogenized by passing through microtip (200ul) 20 x. Collect nuclei by centrifugation at 1000g for 5 min. The nuclei were washed by lysis buffer for 3 times. Resuspended the nuclei in 100ul Tris-EDTA solution (30mM Tris-HCl at pH7.5, 50mM EDTA). Add 1.1 μl concentrated H_2SO_4 and vortex, then incubate on ice more than 1 hour. Take supernatant after centrifugation at 14000rpm for 10 min at 4ºC, and add 1ml acetone to the supernatant. After overnight incubation at RT, centrifuge at 13000rpm for 10 min at RT. Wash the pellet with acetone and air-dry. Resuspend the pellet in water for analysis.

4.2.16 Prepare cell extraction

Cells were resuspended in buffer A (10mM HEPES pH7.9, 5mM $MgCl_2$, 0.25M Sucrose) with 0.1% NP-40. The suspension was passed through an 18G1 needle 4 times and incubated on ice for 10 minutes, then spin at 8000rpm for 10 minutes. The supernatant was kept as cytosolic fraction. The pellet was resuspended in buffer B (10mM HEPES pH7.9, 1mM $MgCl_2$, 0.1mM EDTA, 25% glycerol) with 0.4M NaCl. The suspension was passed through an 18G1 needle 4 times and incubated on ice for 30 minutes, then spin at 10000rpm for 1o minutes. The supernatant was kept as 0.4M nuclear fraction. The pellet was resuspended in buffer B with 2M NaCl. The suspension was passed through an18G1 needle 4 times, incubated on ice for 30 minutes and sonicated 5x 15 seconds on ice. The suspension was kept as 2M nuclear fraction. All the buffers are supplemented by protease inhibitor cocktail (Roche) before use. All fractions of cell extraction were dialyzes against buffer BC100 (25nM Tris-HCl pH7.9, 100mM NaCl, 0.2mM EDTA, 20% glycerol) at 4ºC.

4.2.17 Western blot

Proteins were separated on a SDS-PAGE gel and blotted onto Hybond-P membranes (Amersham). Membranes were probed for primary antibodies and detected with anti-rabbit HRP or anti-mouse HRP secondary antibodies followed by ECL detection (Amersham).

CHAPTER 5. APPENDIX

5.1 Vector information of pGEX-6P-1

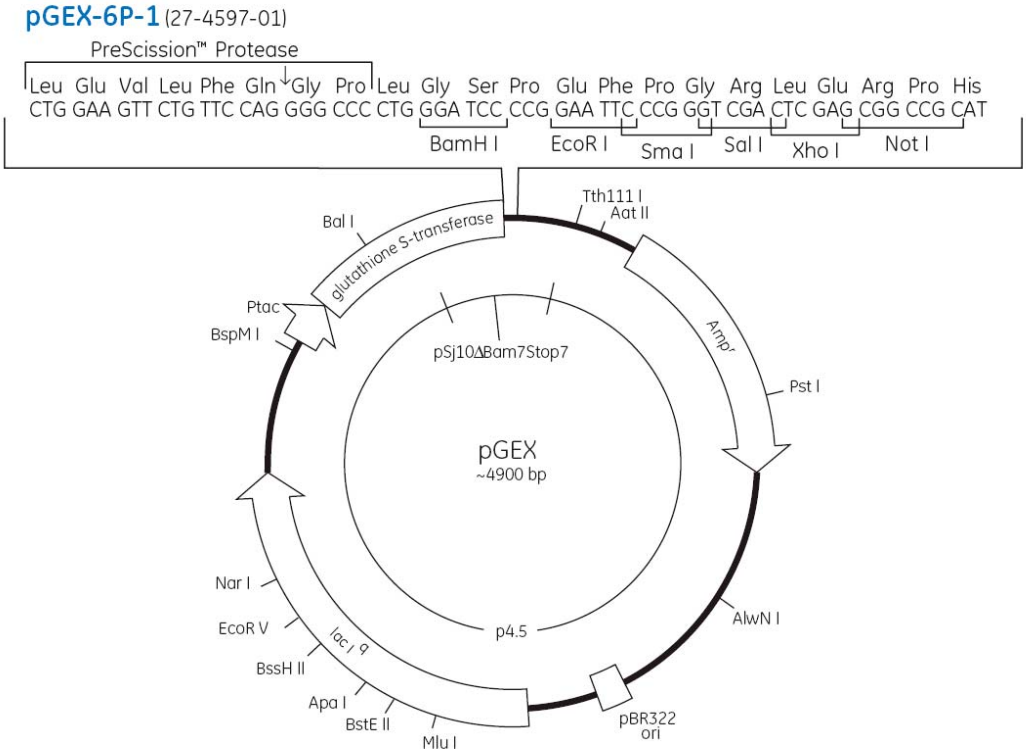

Figure 62. Vector map of pGEX-6P-1

5.2 Vector information of MSCV-LMP

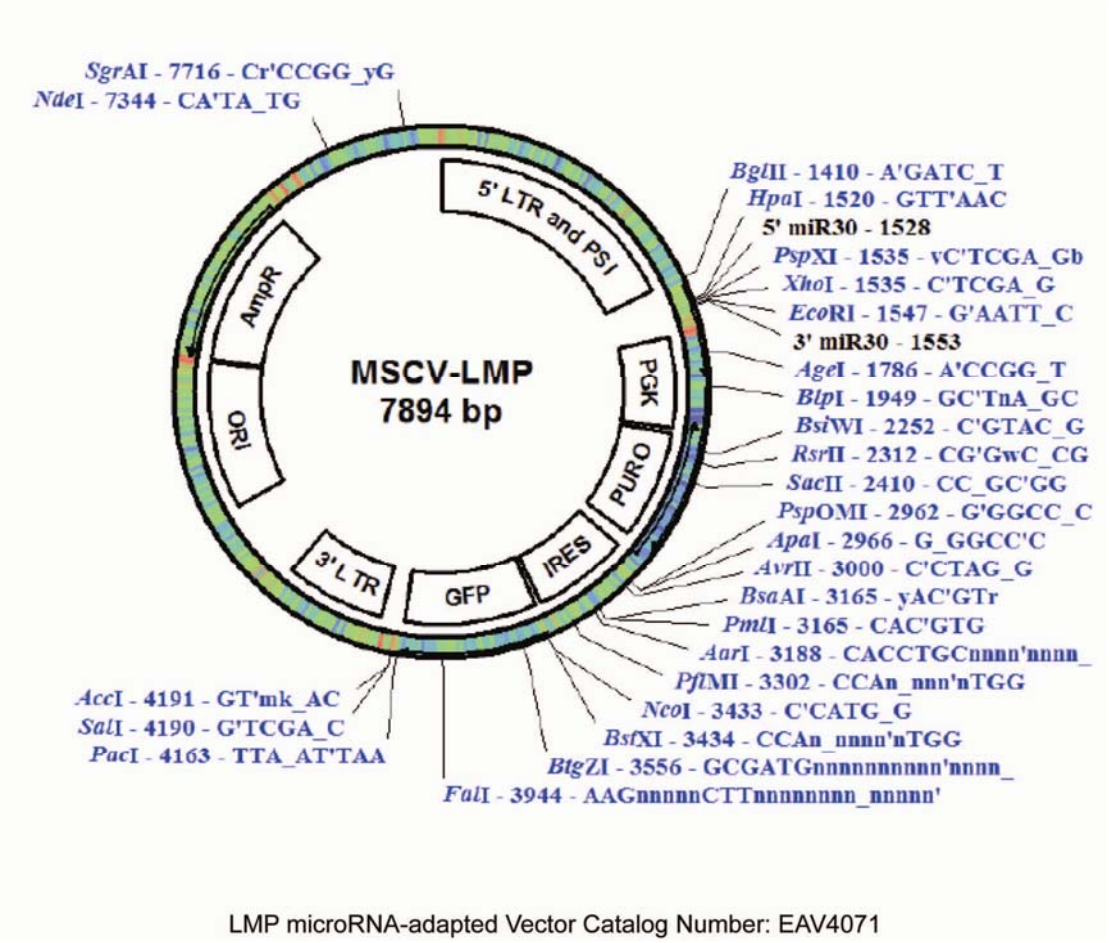

Figure 63. Vector map of MSCV-LMP

5.3 Vector information of pCX-EGFP

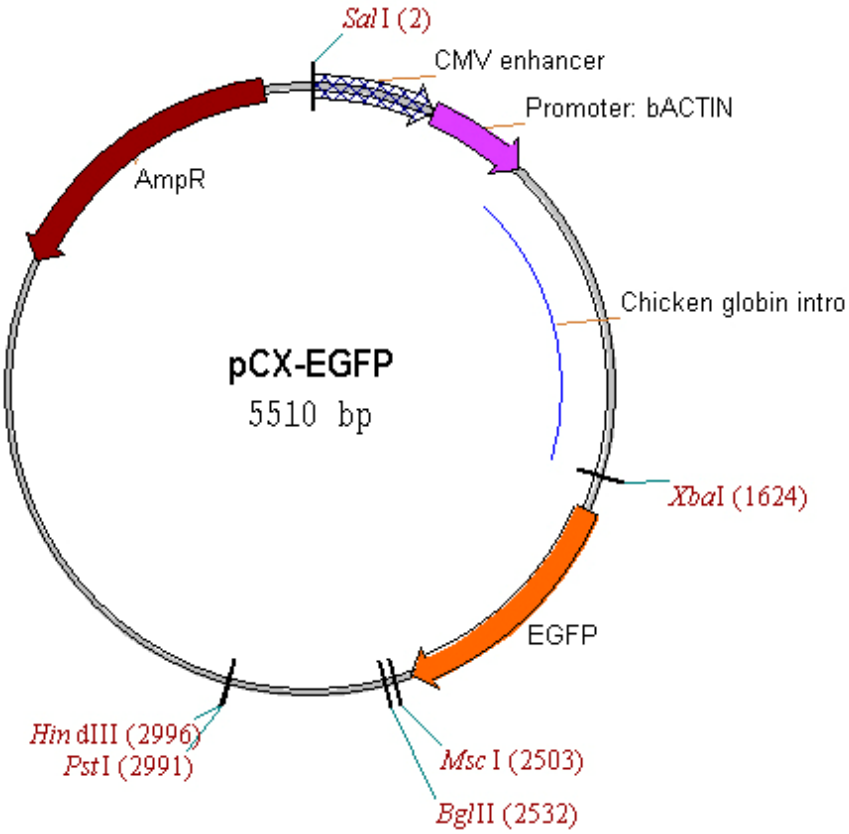

Figure 64. Vector map of pCX-EGFP

REFERENCE:

Adhvaryu, K.K., Morris, S.A., Strahl, B.D., and Selker, E.U. (2005). Methylation of histone H3 lysine 36 is required for normal development in Neurospora crassa. Eukaryotic cell *4*, 1455-1464.

Agger, K., Cloos, P.A., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature.

Allis, C.D., Jenuwein, T., DannyReinberg, and Caparros, M.-L. (2006). Epigenetics (New York, Cold spring harbor laboratory press).

Andersen, E.C., and Horvitz, H.R. (2007). Two C. elegans histone methyltransferases repress lin-3 EGF transcription to inhibit vulval development. Development (Cambridge, England) *134*, 2991-2999.

Arney, K.L., and Fisher, A.G. (2004). Epigenetic aspects of differentiation. Journal of cell science *117*, 4355-4363.

Ayyanathan, K., Lechner, M.S., Bell, P., Maul, G.G., Schultz, D.C., Yamada, Y., Tanaka, K., Torigoe, K., and Rauscher, F.J., III (2003). Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. Genes Dev *17*, 1855-1869.

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.

Barreto, G., Schafer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Doderlein, G., Maltry, N., Wu, W., Lyko, F.*, et al.* (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature *445*, 671-675.

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.

Ben-Shushan, E., Sharir, H., Pikarsky, E., and Bergman, Y. (1995). A dynamic balance between ARP-1/COUP-TFII, EAR-3/COUP-TFI, and retinoic acid receptor:retinoid X receptor heterodimers regulates Oct-3/4 expression in embryonal carcinoma cells. Molecular and cellular biology *15*, 1034-1048.

Bender, L.B., Suh, J., Carroll, C.R., Fong, Y., Fingerman, I.M., Briggs, S.D., Cao, R., Zhang, Y., Reinke, V., and Strome, S. (2006). MES-4: an autosome-associated histone methyltransferase that participates in silencing the X chromosomes in the C. elegans germ line. Development (Cambridge, England) *133*, 3907-3917.

Berger, S.L. (2007). The complex language of chromatin regulation during transcription. Nature *447*, 407-412.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R.*, et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169-181.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. Cell *128*, 669-681.

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., 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. Molecular and cellular biology *26*, 2560-2569.

Biron, V.L., McManus, K.J., Hu, N., Hendzel, M.J., and Underhill, D.A. (2004). Distinct dynamics and distribution of histone methyl-lysine derivatives in mouse development. Developmental biology *276*, 337-351.
Biswas, D., Dutta-Biswas, R., Mitra, D., Shibata, Y., Strahl, B.D., Formosa, T., and Stillman, D.J. (2006). Opposing roles for Set2 and yFACT in regulating TBP binding at promoters. The EMBO journal *25*, 4479-4489.
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.
Brown, M.A., Sims, R.J., 3rd, Gottlieb, P.D., and Tucker, P.W. (2006). Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. Molecular cancer *5*, 26.
Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G., *et al.* (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. Molecular cell *6*, 1287-1295.
Burdon, T., Smith, A., and Savatier, P. (2002). Signalling, cell cycle and pluripotency in embryonic stem cells. Trends in cell biology *12*, 432-438.
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 (New York, NY *298*, 1039-1043.
Cao, R., and Zhang, Y. (2004). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Molecular cell *15*, 57-67.
Cao, S., Bendall, H., Hicks, G.G., Nashabi, A., Sakano, H., Shinkai, Y., Gariglio, M., Oltz, E.M., and Ruley, H.E. (2003). The high-mobility-group box protein SSRP1/T160 is essential for cell viability in day 3.5 mouse embryos. Molecular and cellular biology *23*, 5301-5307.
Carlone, D.L., Lee, J.H., Young, S.R., Dobrota, E., Butler, J.S., Ruiz, J., and Skalnik, D.G. (2005). Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein. Molecular and cellular biology *25*, 4881-4891.
Carrasco, D.R., Tonon, G., Huang, Y., Zhang, Y., Sinha, R., Feng, B., Stewart, J.P., Zhan, F., Khatry, D., Protopopova, M., *et al.* (2006). High-resolution genomic profiles define distinct clinico-pathogenetic subgroups of multiple myeloma patients. Cancer cell *9*, 313-325.
Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., *et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.
Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development (Cambridge, England) *132*, 885-896.
Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.
Chen, T., Tsujimoto, N., and Li, E. (2004). The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Molecular and cellular biology *24*, 9048-9058.
Chew, J.L., Loh, Y.H., Zhang, W., Chen, X., Tam, W.L., Yeap, L.S., Li, P., Ang, Y.S., Lim, B., Robson, P., *et al.* (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Molecular and cellular biology *25*, 6031-6046.

Christensen, J., Agger, K., Cloos, P.A., Pasini, D., Rose, S., Sennels, L., Rappsilber, J., Hansen, K.H., Salcini, A.E., and Helin, K. (2007). RBP2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. Cell *128*, 1063-1076.
Chu, Y., Sutton, A., Sternglanz, R., and Prelich, G. (2006). The BUR1 cyclin-dependent protein kinase is required for the normal pattern of histone methylation by SET2. Molecular and cellular biology *26*, 3029-3038.
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.
Daniel, J.A., Pray-Grant, M.G., and Grant, P.A. (2005). Effector proteins for methylated histones: an expanding family. Cell cycle (Georgetown, Tex *4*, 919-926.
de la Cruz, X., Lois, S., Sanchez-Molina, S., and Martinez-Balbas, M.A. (2005). Do protein motifs read the histone code? Bioessays *27*, 164-175.
Dillon, S.C., Zhang, X., Trievel, R.C., and Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. Genome biology *6*, 227.
Doyon, Y., Cayrou, C., Ullah, M., Landry, A.J., Cote, V., Selleck, W., Lane, W.S., Tan, S., Yang, X.J., and Cote, J. (2006). ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. Molecular cell *21*, 51-64.
Ebert, A., Lein, S., Schotta, G., and Reuter, G. (2006). Histone modification and the control of heterochromatic gene silencing in Drosophila. Chromosome Res *14*, 377-392.
Fang, J., Hogan, G.J., Liang, G., Lieb, J.D., and Zhang, Y. (2007). The Saccharomyces cerevisiae Histone Demethylase Jhd1 Fine-Tunes the Distribution of H3K36me2. Molecular and cellular biology *27*, 5055-5065.
Fehling, H.J., Lacaud, G., Kubo, A., Kennedy, M., Robertson, S., Keller, G., and Kouskoff, V. (2003). Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. Development (Cambridge, England) *130*, 4217-4227.
Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nature cell biology *8*, 188-194.
Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448-453.
Feng, Q., Wang, H., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Struhl, K., and Zhang, Y. (2002). Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. Curr Biol *12*, 1052-1058.
Flanagan, J.F., Mi, L.Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature *438*, 1181-1185.
Flaus, A., and Owen-Hughes, T. (2004). Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer? Current opinion in genetics & development *14*, 165-173.
Francastel, C., Schubeler, D., Martin, D.I., and Groudine, M. (2000). Nuclear compartmentalization and gene activity. Nat Rev Mol Cell Biol *1*, 137-143.
Game, J.C., Williamson, M.S., Spicakova, T., and Brown, J.M. (2006). The RAD6/BRE1 histone modification pathway in Saccharomyces confers radiation resistance through a RAD51-dependent process that is independent of RAD18. Genetics *173*, 1951-1968.
Garvin, C., Holdeman, R., and Strome, S. (1998). The phenotype of mes-2, mes-3, mes-4 and mes-6, maternal-effect genes required for survival of the germline in Caenorhabditis elegans, is sensitive to chromosome dosage. Genetics *148*, 167-185.
Gerber, M., and Shilatifard, A. (2003). Transcriptional elongation by RNA polymerase II and histone methylation. J Biol Chem *278*, 26303-26306.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science (New York, NY *311*, 395-398.
Gu, P., Le Menuet, D., Chung, A.C., and Cooney, A.J. (2006). Differential recruitment of methylated CpG binding domains by the orphan receptor GCNF initiates the repression and silencing of Oct4 expression. Molecular and cellular biology *26*, 9471-9483.
Guenatri, M., Bailly, D., Maison, C., and Almouzni, G. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. The Journal of cell biology *166*, 493-505.
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.
Hao, J., Li, T.G., Qi, X., Zhao, D.F., and Zhao, G.Q. (2006). WNT/beta-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. Developmental biology *290*, 81-91.
Hassan, A.H., Prochasson, P., Neely, K.E., Galasinski, S.C., Chandy, M., Carrozza, M.J., and Workman, J.L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell *111*, 369-379.
Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development (Cambridge, England) *129*, 1983-1993.
Hayakawa, T., Ohtani, Y., Hayakawa, N., Shinmyozu, K., Saito, M., Ishikawa, F., and Nakayama, J. (2007). RBP2 is an MRG15 complex component and down-regulates intragenic histone H3 lysine 4 methylation. Genes Cells *12*, 811-826.
Hebert, J.M., Boyle, M., and Martin, G.R. (1991). mRNA localization studies suggest that murine FGF-5 plays a role in gastrulation. Development (Cambridge, England) *112*, 407-415.
Henikoff, S., Furuyama, T., and Ahmad, K. (2004). Histone variants, nucleosome assembly and epigenetic inheritance. Trends Genet *20*, 320-326.
Hirose, Y., and Ohkuma, Y. (2007). Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. Journal of biochemistry *141*, 601-608.
Hirst, C.E., Ng, E.S., Azzola, L., Voss, A.K., Thomas, T., Stanley, E.G., and Elefanty, A.G. (2006). Transcriptional profiling of mouse and human ES cells identifies SLAIN1, a novel stem cell gene. Developmental biology *293*, 90-103.
Houlard, M., Berlivet, S., Probst, A.V., Quivy, J.P., Hery, P., Almouzni, G., and Gerard, M. (2006). CAF-1 is essential for heterochromatin organization in pluripotent embryonic cells. PLoS genetics *2*, e181.
Huang, Y., Fang, J., Bedford, M.T., Zhang, Y., and Xu, R.M. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science (New York, NY *312*, 748-751.
Huber, T.L., Kouskoff, V., Fehling, H.J., Palis, J., and Keller, G. (2004). Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. Nature *432*, 625-630.
Huyen, Y., Zgheib, O., Ditullio, R.A., Jr., Gorgoulis, V.G., Zacharatos, P., Petty, T.J., Sheston, E.A., Mellert, H.S., Stavridi, E.S., and Halazonetis, T.D. (2004). Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature *432*, 406-411.
Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. (2006). Dissecting self-renewal in stem cells with RNA interference. Nature *442*, 533-538.
Iwai, N., Kitajima, K., Sakai, K., Kimura, T., and Nakano, T. (2001). Alteration of cell adhesion and cell cycle properties of ES cells by an inducible dominant interfering Myb mutant. Oncogene *20*, 1425-1434.

Iwase, S., Lan, F., Bayliss, P., de la Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstine, J.R., Bonni, A., Roberts, T.M., and Shi, Y. (2007). The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. Cell *128*, 1077-1088.
Jorgensen, H.F., Giadrossi, S., Casanova, M., Endoh, M., Koseki, H., Brockdorff, N., and Fisher, A.G. (2006). Stem cells primed for action: polycomb repressive complexes restrain the expression of lineage-specific regulators in embryonic stem cells. Cell cycle (Georgetown, Tex *5*, 1411-1414.
Joshi, A.A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Molecular cell *20*, 971-978.
Kaji, K., Caballero, I.M., MacLeod, R., Nichols, J., Wilson, V.A., and Hendrich, B. (2006). The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nature cell biology *8*, 285-292.
Kaji, K., Nichols, J., and Hendrich, B. (2007). Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. Development (Cambridge, England) *134*, 1123-1132.
Kamakaka, R.T., and Biggins, S. (2005). Histone variants: deviants? Genes & development *19*, 295-310.
Kanno, T., Kanno, Y., Siegel, R.M., Jang, M.K., Lenardo, M.J., and Ozato, K. (2004). Selective recognition of acetylated histones by bromodomain proteins visualized in living cells. Molecular cell *13*, 33-43.
Karachentsev, D., Sarma, K., Reinberg, D., and Steward, R. (2005). PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis. Genes & development *19*, 431-435.
Keogh, M.C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J., *et al.* (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell *123*, 593-605.
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 reports *7*, 397-403.
Kim, T., and Buratowski, S. (2007). Two Saccharomyces cerevisiae JmjC domain proteins demethylate histone H3 K36 in transcribed regions to promote elongation. J Biol Chem.
Kind, J., and Akhtar, A. (2007). Cotranscriptional recruitment of the dosage compensation complex to X-linked target genes. Genes & development *21*, 2030-2040.
Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Molecular and cellular biology *25*, 3305-3316.
Klochendler-Yeivin, A., Fiette, L., Barra, J., Muchardt, C., Babinet, C., and Yaniv, M. (2000). The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression. EMBO reports *1*, 500-506.
Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends in biochemical sciences *31*, 89-97.
Klose, R.J., Gardner, K.E., Liang, G., Erdjument-Bromage, H., Tempst, P., and Zhang, Y. (2007a). Demethylation of histone H3K36 and H3K9 by Rph1: a vestige of an H3K9 methylation system in Saccharomyces cerevisiae? Molecular and cellular biology *27*, 3951-3961.
Klose, R.J., Kallin, E.M., and Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. Nature reviews *7*, 715-727.
Klose, R.J., Yan, Q., Tothova, Z., Yamane, K., Erdjument-Bromage, H., Tempst, P., Gilliland, D.G., Zhang, Y., and Kaelin, W.G., Jr. (2007b). The retinoblastoma binding protein RBP2 is an H3K4 demethylase. Cell *128*, 889-900.

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. PLoS biology *2*, E171.
Kourmouli, N., Jeppesen, P., Mahadevhaiah, S., Burgoyne, P., Wu, R., Gilbert, D.M., Bongiorni, S., Prantera, G., Fanti, L., Pimpinelli, S.*, et al.* (2004). Heterochromatin and tri-methylated lysine 20 of histone H4 in animals. Journal of cell science *117*, 2491-2501.
Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.
Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C.*, et al.* (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Molecular and cellular biology *23*, 4207-4218.
Kurdistani, S.K., and Grunstein, M. (2003). Histone acetylation and deacetylation in yeast. Nat Rev Mol Cell Biol *4*, 276-284.
Kusch, T., Florens, L., Macdonald, W.H., Swanson, S.K., Glaser, R.L., Yates, J.R., 3rd, Abmayr, S.M., Washburn, M.P., and Workman, J.L. (2004). Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions. Science (New York, NY *306*, 2084-2087.
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 & development *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.
Landry, J., Sutton, A., Hesman, T., Min, J., Xu, R.M., Johnston, M., and Sternglanz, R. (2003). Set2-catalyzed methylation of histone H3 represses basal expression of GAL4 in Saccharomyces cerevisiae. Molecular and cellular biology *23*, 5972-5978.
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, M.G., Norman, J., Shilatifard, A., and Shiekhattar, R. (2007a). Physical and functional association of a trimethyl H3K4 demethylase and Ring6a/MBLR, a polycomb-like protein. Cell *128*, 877-887.
Lee, M.G., Villa, R., Trojer, P., Norman, J., Yan, K.P., Reinberg, D., Di Croce, L., and Shiekhattar, R. (2007b). Demethylation of H3K27 Regulates Polycomb Recruitment and H2A Ubiquitination. Science (New York, NY.
Lee, S.H., Oshige, M., Durant, S.T., Rasila, K.K., Williamson, E.A., Ramsey, H., Kwan, L., Nickoloff, J.A., and Hromas, R. (2005). The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair. Proceedings of the National Academy of Sciences of the United States of America *102*, 18075-18080.
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.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.
Leonhardt, H., Page, A.W., Weier, H.U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell *71*, 865-873.
Li, B., Carey, M., and Workman, J.L. (2007a). The role of chromatin during transcription. Cell *128*, 707-719.
Li, B., Gogol, M., Carey, M., Lee, D., Seidel, C., and Workman, J.L. (2007b). Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. Science (New York, NY *316*, 1050-1054.
Li, B., Gogol, M., Carey, M., Pattenden, S.G., Seidel, C., and Workman, J.L. (2007c). Infrequently transcribed long genes depend on the Set2/Rpd3S pathway for accurate transcription. Genes & development *21*, 1422-1430.

Li, B., Howe, L., Anderson, S., Yates, J.R., 3rd, and Workman, J.L. (2003). The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem *278*, 8897-8903.
Liang, G., Chan, M.F., Tomigahara, Y., Tsai, Y.C., Gonzales, F.A., Li, E., Laird, P.W., and Jones, P.A. (2002). Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Molecular and cellular biology *22*, 480-491.
Liang, G., Klose, R.J., Gardner, K.E., and Zhang, Y. (2007). Yeast Jhd2p is a histone H3 Lys4 trimethyl demethylase. Nature structural & molecular biology *14*, 243-245.
Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods (San Diego, Calif *25*, 402-408.
Lowell, S., Benchoua, A., Heavey, B., and Smith, A.G. (2006). Notch promotes neural lineage entry by pluripotent embryonic stem cells. PLoS biology *4*, e121.
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.
Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. The EMBO journal *24*, 800-812.
Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol *6*, 838-849.
Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., *et al.* (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nature cell biology *9*, 625-635.
Mateescu, B., England, P., Halgand, F., Yaniv, M., and Muchardt, C. (2004). Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. EMBO reports *5*, 490-496.
Mellor, J. (2006a). Dynamic nucleosomes and gene transcription. Trends Genet *22*, 320-329.
Mellor, J. (2006b). It takes a PHD to read the histone code. Cell *126*, 22-24.
Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol *7*, 540-546.
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. Developmental cell *10*, 105-116.
Miao, F., and Natarajan, R. (2005). Mapping global histone methylation patterns in the coding regions of human genes. Molecular and cellular biology *25*, 4650-4661.
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.
Minucci, S., Botquin, V., Yeom, Y.I., Dey, A., Sylvester, I., Zand, D.J., Ohbo, K., Ozato, K., and Scholer, H.R. (1996). Retinoic acid-mediated down-regulation of Oct3/4 coincides with the loss of promoter occupancy in vivo. The EMBO journal *15*, 888-899.
Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.
Montgomery, N.D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S.J., Otte, A.P., and Magnuson, T. (2005). The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. Curr Biol *15*, 942-947.
Morillon, A., Karabetsou, N., Nair, A., and Mellor, J. (2005). Dynamic lysine methylation on histone H3 defines the regulatory phase of gene transcription. Molecular cell *18*, 723-734.

Morris, S.A., Rao, B., Garcia, B.A., Hake, S.B., Diaz, R.L., Shabanowitz, J., Hunt, D.F., Allis, C.D., Lieb, J.D., and Strahl, B.D. (2007). Identification of histone H3 lysine 36 acetylation as a highly conserved histone modification. J Biol Chem *282*, 7632-7640.
Morris, S.A., Shibata, Y., Noma, K., Tsukamoto, Y., Warren, E., Temple, B., Grewal, S.I., and Strahl, B.D. (2005). Histone H3 K36 methylation is associated with transcription elongation in Schizosaccharomyces pombe. Eukaryotic cell *4*, 1446-1454.
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.
Muller, W.G., Rieder, D., Karpova, T.S., John, S., Trajanoski, Z., and McNally, J.G. (2007). Organization of chromatin and histone modifications at a transcription site. The Journal of cell biology *177*, 957-967.
Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science (New York, NY *292*, 110-113.
Ng, E.S., Azzola, L., Sourris, K., Robb, L., Stanley, E.G., and Elefanty, A.G. (2005). The primitive streak gene Mixl1 is required for efficient haematopoiesis and BMP4-induced ventral mesoderm patterning in differentiating ES cells. Development (Cambridge, England) *132*, 873-884.
Nicolas, E., Yamada, T., Cam, H.P., Fitzgerald, P.C., Kobayashi, R., and Grewal, S.I. (2007). Distinct roles of HDAC complexes in promoter silencing, antisense suppression and DNA damage protection. Nature structural & molecular biology *14*, 372-380.
Nishimoto, M., Miyagi, S., Yamagishi, T., Sakaguchi, T., Niwa, H., Muramatsu, M., and Okuda, A. (2005). Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus. Molecular and cellular biology *25*, 5084-5094.
Nishioka, K., Rice, J.C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R., *et al.* (2002). PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. Molecular cell *9*, 1201-1213.
Niwa, H. (2007). How is pluripotency determined and maintained? Development (Cambridge, England) *134*, 635-646.
Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nature genetics *24*, 372-376.
Noma, K., Allis, C.D., and Grewal, S.I. (2001). Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science (New York, NY *293*, 1150-1155.
Ohbayashi, T., Shimada, M., Nakadai, T., Wada, T., Handa, H., and Tamura, T. (2003). Vertebrate TBP-like protein (TLP/TRF2/TLF) stimulates TATA-less terminal deoxynucleotidyl transferase promoters in a transient reporter assay, and TFIIA-binding capacity of TLP is required for this function. Nucleic acids research *31*, 2127-2133.
Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V.M., Su, L., Xu, G., and Zhang, Y. (2005). hDOT1L links histone methylation to leukemogenesis. Cell *121*, 167-178.
Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.
Okuda, A., Fukushima, A., Nishimoto, M., Orimo, A., Yamagishi, T., Nabeshima, Y., Kuro-o, M., Nabeshima, Y., Boon, K., Keaveney, M., *et al.* (1998). UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. The EMBO journal *17*, 2019-2032.
Ouaissi, M., and Ouaissi, A. (2006). Histone deacetylase enzymes as potential drug targets in cancer and parasitic diseases. Journal of biomedicine & biotechnology *2006*, 13474.

Ovitt, C.E., and Scholer, H.R. (1998). The molecular biology of Oct-4 in the early mouse embryo. Molecular human reproduction *4*, 1021-1031.
Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Molecular and cellular biology *27*, 3769-3779.
Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y.*, et al.* (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Molecular cell *12*, 1577-1589.
Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E.*, et al.* (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517-527.
Polioudaki, H., Markaki, Y., Kourmouli, N., Dialynas, G., Theodoropoulos, P.A., Singh, P.B., and Georgatos, S.D. (2004). Mitotic phosphorylation of histone H3 at threonine 3. FEBS letters *560*, 39-44.
Pray-Grant, M.G., Daniel, J.A., Schieltz, D., Yates, J.R., 3rd, and Grant, P.A. (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature *433*, 434-438.
Prigent, C., and Dimitrov, S. (2003). Phosphorylation of serine 10 in histone H3, what for? Journal of cell science *116*, 3677-3685.
Pugh, B.F., and Tjian, R. (1991). Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes & development *5*, 1935-1945.
Rayasam, G.V., Wendling, O., Angrand, P.O., Mark, M., Niederreither, K., Song, L., Lerouge, T., Hager, G.L., Chambon, P., and Losson, R. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. The EMBO journal *22*, 3153-3163.
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.
Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annual review of genetics *38*, 413-443.
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. Molecular cell *25*, 15-30.
Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. Nat Rev Mol Cell Biol *7*, 437-447.
Sakaki-Yumoto, M., Kobayashi, C., Sato, A., Fujimura, S., Matsumoto, Y., Takasato, M., Kodama, T., Aburatani, H., Asashima, M., Yoshida, N.*, et al.* (2006). The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. Development (Cambridge, England) *133*, 3005-3013.
Sanders, S.L., Portoso, M., Mata, J., Bahler, J., Allshire, R.C., and Kouzarides, T. (2004). Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. Cell *119*, 603-614.
Sauter, C.N., McDermid, R.L., Weinberg, A.L., Greco, T.L., Xu, X., Murdoch, F.E., and Fritsch, M.K. (2005). Differentiation of murine embryonic stem cells induces progesterone receptor gene expression. Experimental cell research *311*, 251-264.
Schneider, R., Bannister, A.J., and Kouzarides, T. (2002). Unsafe SETs: histone lysine methyltransferases and cancer. Trends in biochemical sciences *27*, 396-402.
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. Nature cell biology *6*, 73-77.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes & development *18*, 1251-1262.
Schubeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., Delrow, J.*, et al.* (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes & development *18*, 1263-1271.
Shanower, G.A., Muller, M., Blanton, J.L., Honti, V., Gyurkovics, H., and Schedl, P. (2005). Characterization of the grappa gene, the Drosophila histone H3 lysine 79 methyltransferase. Genetics *169*, 173-184.
Shen, M.M., and Leder, P. (1992). Leukemia inhibitory factor is expressed by the preimplantation uterus and selectively blocks primitive ectoderm formation in vitro. Proceedings of the National Academy of Sciences of the United States of America *89*, 8240-8244.
Shi, X., Kachirskaia, I., Walter, K.L., Kuo, J.H., Lake, A., Davrazou, F., Chan, S.M., Martin, D.G., Fingerman, I.M., Briggs, S.D.*, et al.* (2007). Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. J Biol Chem *282*, 2450-2455.
Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.
Shi, Y., and Whetstine, J.R. (2007). Dynamic regulation of histone lysine methylation by demethylases. Molecular cell *25*, 1-14.
Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science (New York, NY *311*, 844-847.
Smallwood, A., Esteve, P.-O., Pradhan, S., and Carey, M. (2007). Functional cooperation between HP1 and DNMT1 mediates gene silencing. Genes Dev *21*, 1169-1178.
Stabell, M., Larsson, J., Aalen, R.B., and Lambertsson, A. (2007). Drosophila dSet2 functions in H3-K36 methylation and is required for development. Biochemical and biophysical research communications *359*, 784-789.
Sterner, D.E., and Berger, S.L. (2000). Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev *64*, 435-459.
Stopka, T., and Skoultchi, A.I. (2003). The ISWI ATPase Snf2h is required for early mouse development. Proceedings of the National Academy of Sciences of the United States of America *100*, 14097-14102.
Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.
Strahl, B.D., Grant, P.A., Briggs, S.D., Sun, Z.W., Bone, J.R., Caldwell, J.A., Mollah, S., Cook, R.G., Shabanowitz, J., Hunt, D.F.*, et al.* (2002). Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Molecular and cellular biology *22*, 1298-1306.
Sun, X.J., Wei, J., Wu, X.Y., Hu, M., Wang, L., Wang, H.H., Zhang, Q.H., Chen, S.J., Huang, Q.H., and Chen, Z. (2005). Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. J Biol Chem *280*, 35261-35271.
Takahashi, K., Murakami, M., and Yamanaka, S. (2005). Role of the phosphoinositide 3-kinase pathway in mouse embryonic stem (ES) cells. Biochemical Society transactions *33*, 1522-1525.
Talasz, H., Lindner, H.H., Sarg, B., and Helliger, W. (2005). Histone H4-lysine 20 monomethylation is increased in promoter and coding regions of active genes and correlates with hyperacetylation. J Biol Chem *280*, 38814-38822.

Tanaka, Y., Katagiri, Z.-i., Kawahashi, K., Kioussis, D., and Kitajima, S. (2007). Trithorax-group protein ASH1 methylates histone H3 lysine 36. gene
397, 9.
Tremethick, D.J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. Cell *128*, 651-654.
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.
Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.
Tu, S., Bulloch, E.M., Yang, L., Ren, C., Huang, W.C., Hsu, P.H., Chen, C.H., Liao, C.L., Yu, H.M., Lo, W.S., *et al.* (2007). Identification of histone demethylases in Saccharomyces cerevisiae. J Biol Chem *282*, 14262-14271.
Turner, B.M. (2000). Histone acetylation and an epigenetic code. Bioessays *22*, 836-845.
Vakoc, C.R., Mandat, S.A., Olenchock, B.A., and Blobel, G.A. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Molecular cell *19*, 381-391.
Vakoc, C.R., Sachdeva, M.M., Wang, H., and Blobel, G.A. (2006). Profile of histone lysine methylation across transcribed mammalian chromatin. Molecular and cellular biology *26*, 9185-9195.
van Leeuwen, F., Gafken, P.R., and Gottschling, D.E. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell *109*, 745-756.
Vaquero, A., Scher, M.B., Lee, D.H., Sutton, A., Cheng, H.L., Alt, F.W., Serrano, L., Sternglanz, R., and Reinberg, D. (2006). SirT2 is a histone deacetylase with preference for histone H4 Lys 16 during mitosis. Genes & development *20*, 1256-1261.
Wang, A., Kurdistani, S.K., and Grunstein, M. (2002a). Requirement of Hos2 histone deacetylase for gene activity in yeast. Science (New York, NY *298*, 1412-1414.
Wang, B., Matsuoka, S., Carpenter, P.B., and Elledge, S.J. (2002b). 53BP1, a mediator of the DNA damage checkpoint. Science (New York, NY *298*, 1435-1438.
Wang, G.G., Cai, L., Pasillas, M.P., and Kamps, M.P. (2007). NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. Nature cell biology *9*, 804-812.
Wang, J., Mager, J., Chen, Y., Schneider, E., Cross, J.C., Nagy, A., and Magnuson, T. (2001). Imprinted X inactivation maintained by a mouse Polycomb group gene. Nature genetics *28*, 371-375.
Weber, M., and Schubeler, D. (2007). Genomic patterns of DNA methylation: targets and function of an epigenetic mark. Current opinion in cell biology *19*, 273-280.
Weis, L., and Reinberg, D. (1997). Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. Molecular and cellular biology *17*, 2973-2984.
Whetstine, J.R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M., *et al.* (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell *125*, 467-481.
Wu, J., Suka, N., Carlson, M., and Grunstein, M. (2001). TUP1 utilizes histone H3/H2B-specific HDA1 deacetylase to repress gene activity in yeast. Molecular cell *7*, 117-126.
Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., *et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86-90.

Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., and Strahl, B.D. (2003). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes & development *17*, 654-663.
Xu, F., Zhang, K., and Grunstein, M. (2005). Acetylation in histone H3 globular domain regulates gene expression in yeast. Cell *121*, 375-385.
Yamane, K., Tateishi, K., Klose, R.J., Fang, J., Fabrizio, L.A., Erdjument-Bromage, H., Taylor-Papadimitriou, J., Tempst, P., and Zhang, Y. (2007). PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. Molecular cell *25*, 801-812.
Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development (Cambridge, England) *122*, 881-894.
Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281-292.
Zhang, J., Tam, W.L., Tong, G.Q., Wu, Q., Chan, H.Y., Soh, B.S., Lou, Y., Yang, J., Ma, Y., Chai, L.*, et al.* (2006a). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. Nature cell biology *8*, 1114-1123.
Zhang, P., Du, J., Sun, B., Dong, X., Xu, G., Zhou, J., Huang, Q., Liu, Q., Hao, Q., and Ding, J. (2006b). Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. Nucleic acids research *34*, 6621-6628.
Zhao, Z., Yu, Y., Meyer, D., Wu, C., and Shen, W.H. (2005). Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. Nature cell biology *7*, 1256-1260.

Acknowledgements

I thank Dr. Fang-Lin Sun, who gave me the opportunity to start my PhD study at the FMI. I highly appreciate Dr. Antoine Peters, who took me over in his lab to finish my PhD study after Dr. Sun had left the FMI. Dr. Peters showed me how a scientist should be and helped me very much to finish my thesis work with his understanding, his clear vision, his great knowledge, and his enthusiasm.

I thank Prof. Susan Gasser, Prof. Primo Schär, and Dr. Dirk Schübeler to be my thesis committee members, as well as Prof. Denis Monard and Prof. Nancy Hynes.

I thank Fred for his great help with the HMT assay and Western blot analysis during ES cell differentiation. I thank Mareike for her beautiful pictures of the H3K36 methylation stainings in the early embryo. I thank Philip for the many good discussions. I thank also all the other members of the Peters lab for their kindly help and friendly communication. All support I received in this lab made my life during the last two years at the FMI easier and happier.

I thank all my colleagues in my former lab. Especially, Vanda and Dejing helped me a lot. I was lucky to be able to share my pleasure and to relief my stress with them during the beginning days in Basel.

I thank all the other people at FMI who have ever helped me.

Finally, great thanks go to my family, especially my parents as well as my wife Chun. Their love for me, every moment during the past years, has been really important for me to finish my PhD study.

Curriculum Vitae

Name: Tianke Wang
Date of Birth: October 07th, 1975
Nationality: China
Address: Friedrich Miescher Institute for Biomedical Research
Part of the Novartis Research Foundation
Maulbeerstrasse 66, CH-4058, Basel
Telephone: 061 6976 454
076 5785 109
E-Mail: tianke@fmi.ch
tiankewang@yahoo.com

My Personal Highlights:

- Highly motivated team-player, well-organized, and with profound knowledge on biomedical research.
- Five years research experience in epigenetics in an international group in Switzerland.
- Three years experience in process development of antibiotic fermentation and practical experience in production of antibiotics in pharmaceutical industry.
- Responsibilities in social/extracurricular communities.

Education:

Mar. 2002 - present Ph. D
Epigenetics Department, Friedrich Miescher Institute, Basel, Switzerland

Sep. 1998 - Jul. 2001 M.S.
Department of Microbiology, China Agricultural University, Beijing, China

Sep. 1994 - Jul. 1998 B.S.
Department of Microbiology, China Agricultural University, Beijing, China

Professional Experience:

Mar. 2002 - present Ph. D student
Epigenetics Department, Friedrich Miescher Institute, Basel, Switzerland

- Investigated biological function of Histone H3 lysine 36 methylation in embryonic stem cell self renewal and differentiation

Nsd proteins and Hypb are putative histone H3 lysine 36 methyltransferases that have been implicated in different human diseases. I investigated the enzymatic activity of these proteins by both *in vitro* assay and *in vivo* cellular system. I analyzed the function of H3 lysine 36 methylation during embryonic stem cell differentiation.

Jul. 2001 – Feb. 2002 Research/Teaching Assistant
Department of Microbiology, China Agricultural University (CAU), Beijing, China

- Screened avermectin high yield strains and studied genetic instability of *Streptomyces avermitilis*
- Instructed experimental courses of Microbiology for undergraduate students

Aug. 1999 - Feb. 2002 Technical Supporter
Qilu Pharmaceutical Co., LTD, Shandong, China

- Finished laboratory process development

- Provided technical support from pilot scale fermentation to manufacture of avermectin

Sep. 1998 - Jun. 2001 Postgraduate
Department of Microbiology, China Agricultural University, Beijing, China

- Established a new method for screening of avermectin high yield strains

I established a simple and rapid method to screen avermectin yield strain. This reduced material costs and shortened the screening and fermentation cycle.

- Improved production yield of avermectin in manufacturing

Avermectin is a family of macrolide antibiotic against parasites that is widely applied in agriculture and medicine. It is one of the key products of Qilu Pharmaceutical Co., LTD. With my research and technical support, I contributed to improve the yield of avermectin and brought Qilu Pharmaceutical Co., LTD a considerable commercial profit.

Sep. 1997 - Aug.1998 Trainee
Department of Microbiology, China Agricultural University, Beijing, China

- Studied the differentiation of *Streptomyces avermitilis*

Genetic instability of *Streptomyces avermitilis* affects the yield of avermectin for industry production. I characterized the different stains from spontaneous differentiation by morphology and initiated the project to understand the underlying molecular mechanism of genetic instability.

Techniques Acquired:

- **Cell biology:** Mammalian cell culture, Insect cell culture, Transfection, Retroviral infection, Stable cell line establishment, Recombinant protein expression, Cell cycle analysis, Cell proliferation and apoptosis analysis, FACS, Immunofluorescence applications
- **Molecular biology:** Molecular cloning, Western blot, RNAi, quantitative RT-PCR, Chromatin Immunoprecipitation, Microarray analysis
- **Biochemistry:** Immunoprecipitation, protein expression and purification, HPLC
- **Animal work:** Mouse handling and organization of large mouse colonies
- **Microbiology:** Aseptic technique, Isolation and pure culture of bacterial/fungi, Microscopy and stains, Classification of bacteria/fungi, Mutagenesis, Fermentation technology
- **Microscopy:** Fluorescence microscopy and confocal microscopy, Imaging software

Certificate/Workshop:

2002	Radiation protection course
2004	Module I: introductory course in laboratory animal science,
	Friedrich Miescher Institute, Basel, Switzerland
2005	Scientific and Technical Writing, MediWrite, Basel, Switzerland

Extracurricular Activities:

1999	President of graduate council of College of Biological Sciences, CAU
1996-1997	President of undergraduate council of College of Biological Sciences
1995-1998	Member of the basketball team of College of Biological Sciences

1995	Chief Editor of College Newspaper “Xinyu”

Awards and Honors:

1998	Outstanding Graduate of all Beijing
1998	Best Thesis of Department of Microbiology, CAU
1995-1997	Successive 3 years University Scholarship
1996	University excellent leader of the student council

Tianke Wang

Schönaustrasse 78

4058 Basel

den 18. September 2007

Herrn

Prof. Dr. Hans-Peter Hauri

Universität Basel

Dekanat der der Philosophisch-

Naturwissenschaftlichen Fakultät

Klingelbergstrasse 50
4056 Basel

Betreff: Zulassung zum Doktoratsexamen

Sehr geehrter Herr Professor Hauri,

Hiermit möchte ich die Dissertationsarbeit zur Zulassung zum Doktoratsexamen vorlegen.

Ich erkläre, dass ich die Dissertation „The Histone H3 Lysine 36 Methyltransferse HypB Facilitates Embryonic Stem Cell differentiation“ nur mit der darin angegebenen Hilfe verfasst und bei keiner anderen Universität und keiner anderen Fakultät der Universität Basel eingereicht habe.

Beigelegt ist die Dissertation mit meinem Lebenslauf.

Mit freundlichen Empfehlungen

Tianke Wang