

Nanog-regulated Histone Demethylation in Embryonic Stem Cell Identity

**By
Liangqi Xie**

A DISSERTATION

Presented to the Department of
Cell & Developmental Biology
and the Oregon Health & Science University
School of Medicine
in partial fulfillment of
the requirement for the degree of
Doctor of Philosophy
April 2012

School of Medicine
Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. dissertation of
Liangqi Xie
has been approved

Mentor/Advisor : Soren Impey, Ph.D.

Committee Chair : Rosalie Sears, Ph.D.

Member : Richard Goodman, M.D., Ph.D.

Member : Markus Grompe, M.D.

Member : Mathew Thayer, Ph.D.

TABLE OF CONTENTS

Table of Contents	ii
List of Figures and Tables	iv
List of Abbreviations	vi
Acknowledgements	viii
Abstract	x
Chapter One: Introduction	1
A Brief History of Embryonic Stem Cell Research	2
Pluripotent Stem Cell Characteristics.....	4
Signaling Pathways that Regulate Pluripotent Stem Cells	8
Genetic Network of Core Pluripotency Transcription Factors.....	10
Epigenetic Mechanisms Governing Pluripotent Stem Cells.....	13
Nanog Connects Genetic and Epigenetic Programs for ESC Identity	15
KDM5 Family of Histone Demethylases.....	21
Chapter Two: KDM5B Regulates Embryonic Stem Cell Self-Renewal by Repressing Cryptic Intragenic Transcription	34
Abstract	36
Introduction.....	37
Results	39
Discussion.....	48
Acknowledgements	53
Figure and Figure legends.....	54
Methods	76

Chapter Three: KDM5B, an Epigenetic Module Regulating Embryonic Stem Cell Pluripotency.....	81
Abstract	83
Introduction.....	84
Results	86
Discussion.....	97
Acknowledgements	101
Figure and Figure legends.....	102
Methods	126
Chapter Four: Summary, Discussion and Conclusion	133
Summary.....	134
Discussion.....	139
Conclusion.....	155
Appendices	156
Figures	156
Contribution to Figures	161
References.....	162

LIST OF FIGURES AND TABLES

Figure 1.1: Origin of pluripotent stem cells at different stages of the mouse embryonic development	5
Figure 1.2: A complex signaling network regulating the pluripotent state	8
Figure 1.3: Flow chart of functional genomics approaches to interrogate Nanog-regulated epigenetic mechanisms towards ESC naïve pluripotency.....	17
Figure 1.4: Biochemical mechanisms of LSD1-mediated histone demethylation.....	21
Figure 1.5: Biochemical mechanisms of JmjC domain-containing histone lysine demethylases (KDMs)	22
Figure 1.6: Domain architecture comparison of KDM5/JARID family histone demethylase proteins.....	23
Figure 1.7: De novo generation of phylogenetic relationships among KDM5/JARID1 family members from model organisms	24
Figure 1.8: Structural analysis of the catalytic JmjC domain	26
Table 1.1: List of Nanog-regulated chromatin modulator genes	19
Table 1.2: A comparison between Kdm5b and Nanog homozygous mutant embryos...31	
Figure 2.1: <i>Kdm5b</i> is a Nanog and Oct4 target and critical for ESC self-renewal.....	54
Figure 2.2: An unexpected role for KDM5B in transcriptional activation	56
Figure 2.3: KDM5B removes local domains of intragenic H3K4me3	58
Figure 2.4: H3K36me3 recruits KDM5B to transcriptionally active intragenic regions ...	60
Figure 2.5: MRG15 mediates recruitment of KDM5B	62
Figure 2.6: KDM5B safeguards transcriptional elongation by repressing cryptic transcription.....	64
Supplemental Figure 2.1-2.10:.....	66-75
Figure 3.1: Functional genomics assays identify novel Nanog-regulated chromatin modulators regulating ESC identity	102
Figure 3.2: <i>Kdm5b</i> is highly enriched in pluripotent cells and tissues.....	104
Figure 3.3: <i>Kdm5b</i> expression maintains ESC self-renewal in the absence of LIF.....	106

Figure 3.4: <i>Kdm5b</i> supports ESCs pluripotency following long-term LIF-withdrawal culture.....	108
Figure 3.5: <i>Kdm5b</i> is downstream effector of Nanog	110
Figure 3.6: Cell cycle and DNA synthesis pathways mediate KDM5B-regulated ESC self-renewal	112
Figure 3.7: Characterizing the KDM5B demethylase complex	114
Supplemental Figure 3.1-3.5:	116-125
Figure 4.1: Model of KDM5B in regulating naïve pluripotent state.....	155
Appendix Figure 1	156
Appendix Figure 2	157
Appendix Figure 3	158
Appendix Figure 4	159
Appendix Figure 5.....	160

LIST OF ABBREVIATIONS

AML	Acute myeloid leukemia
AP	Alkaline phosphatase
ChIP	Chromatin immunoprecipitation
ChIP-Seq	Chromatin immunoprecipitation and high-throughput sequencing
DRB	5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole
E5B	E14Tg2a ESCs stably expressing Kdm5b transgene
ECC	Embryonic carcinoma cell
EGC	Embryonic germ cell
EpiSC	Epiblast stem cell
ESC	Embryonic stem cell
FAD	Reduced flavin adenine dinucleotide
GRO-Seq	Genome-wide nuclear run-on and high-throughput sequencing
H3K4me3	Histone 3 Lysine 4 tri-methylation
H3K4me36	Histone 3 Lysine 36 tri-methylation
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HMT	Histone methyltransferase
ICM	Inner cell mass
iPSC	Induced pluripotent stem cell
JHDM	JmjC-domain-containing histone demethylases
JmjC	Jumonji C

Kdm5b	Lysine demethylase 5b
Lid	Little imaginal discs
LIF	Leukemia inhibitory factor
MAPK	Mitogen-activated protein kinase
MO	Morpholino
NSC	Neural stem cell
PcG	Polycomb group protein
PE	Primitive endoderm
PGC	Primordial germ cell
PI3K	Phosphoinositide 3-kinase
P-TEFb	Positive transcription elongation factor b
RA	Retinoic acid
RNA-Seq	RNA high-throughput sequencing
shRNA	Short Hairpin RNA
siRNA	Small Interfering RNA
Stat3	Signal transducer and activator of transcription-3
TE	Trophectoderm
TF	Transcription factor
TrxG	Trithorax group protein
TSA	Trichostatin A
α KG	α -ketoglutarate

ACKNOWLEDGEMENTS

First, I would like to send my most sincere gratitude to my thesis mentor—Dr. Soren Impey. I still clearly remember his provocative enthusiasm to science when I first talked with him about a potential rotation possibility. His insight on studying the burgeoning field combining the stem cell biology, epigenetic mechanisms and bioinformatics is another reason prompting me to join his lab. Throughout my graduate study, Soren's scientific passion and wisdom never fade. My research progress will not be possible without Soren's sharp insights on finding the right way resolving a problem, calm tolerance on my experimental failures, patient back-and-forth on revising the manuscripts, warm generosity in concentrating resources on my projects, and motivated encouragement when I face experimental frustrations. Thinking big and looking fundamental characterizes his scientific style and taste, which have deeply motivated me and will continue to penetrate into my mind throughout my scientific career.

I would extend my gratitude to Dr. Markus Grompe, who has founded a very collaborative research environment in the Oregon Stem Cell Center and provided me many experimental help and guidance. Working in such a beautiful and creative environment is absolutely one of the most joyful moments in my life. I would also thank Dr. Rosie Sears and Dr. Richard Goodman for their valuable insights on my research projects and many career-wise suggestions and support.

I will also remember the many assistance and help from all the former and current members in the Soren lab. Carl Pelz's strong computational ability has helped

shed the first insight into Kdm5b's molecular mechanisms on transcriptional elongation. Sean Shadle's smart hands have contributed significantly to the progress of the Tet2 project. And my progress would not be separable from the hard work and friendly coordination from Wensi Wang, Amir Bashar, Olga Varlamova, Fan Yang and Hideaki Ando.

Finally, I would like to send my special thanks to my parents Cuifang Wang and Jinming Xie for being my constant support and my sole mate, Zhiling Zhang, who has accompanied me many late nights in the lab and has enlightened me to gradually understand the meaning of family, duty, and life.

ABSTRACT

Pluripotent embryonic stem cells (ESCs) are derived from the epiblast of pre-implantation embryos and have two defining features: indefinite self-renewal *in vitro* and pluripotency-- the capacity to differentiate into all somatic cells and the germ lineage. ESCs depend on the leukemia inhibitory factor (LIF) signaling pathways to maintain the naïve pluripotent state and therefore can readily incorporate the inner cell mass (ICM) to resume normal embryonic development. Understanding the molecular mechanisms that establish and maintain the ESCs pluripotency is essential for their application to regenerative medicine. The core pluripotency factor Nanog is proposed to specify and stabilize a naïve pluripotent state in both embryogenesis and induced reprogramming. Nevertheless, the downstream epigenetic mechanisms by which Nanog constructs an epigenetic landscape permissive to the naïve pluripotent state are unclear.

Nanog ChIP-Seq and Nanog knockdown RNA-Seq analysis identify histone H3 trimethyl lysine 4 (H3K4me3) demethylase *Kdm5b* as a direct downstream target of Nanog in mouse ESCs. *Kdm5b* expression is highly enriched in the ICM of mouse pre-implantation embryos and pluripotency-associated cells and tissues. Acute *Kdm5b* mRNA knockdown impairs ESCs proliferation and self-renewal. Forced expression of *Kdm5b* can sustain ESCs in an undifferentiated state in the absence of LIF signaling for up to two months. These cells are still pluripotent because they can contribute to teratoma formation. Epistasis experiments demonstrate that *Kdm5b* overexpression can compensate for loss of Nanog but not Klf4, suggesting that *Kdm5b* is a major downstream effector of Nanog. Ectopic expression of *Kdm5b* can also boost the reprogramming efficiency from neural stem cells in conjunction with Oct4 or Oct4/Klf4.

These lines of evidence suggest that *Kdm5b* is an important epigenetic regulator for the ESC naïve pluripotent state.

Although KDM5B is believed to function as a promoter-bound repressor, analysis of *Kdm5b* knockdown RNA-Seq suggests that it paradoxically functions as an activator of a gene network associated with ESC self-renewal. Our ChIP-Seq data reveals that KDM5B predominantly occupies active intragenic regions and significantly correlates with the elongating RNA Pol II and histone H3 trimethyl lysine 36 (H3K36me3). KDM5B is recruited to H3K36me3 via an interaction with the chromodomain protein MRG15. Genome-wide analysis of MRG15 occupancy shows a high degree of colocalization with KDM5B. *Mrg15* is an ortholog of *eaf3*, a component of the yeast Rpd3S complex which functions to repress cryptic intragenic transcription. Likewise, knockdown of *Kdm5b* or *Mrg15* significantly increases cryptic intragenic transcription, promotes unphosphorylated RNA Pol II recruitment and reduces elongation-associated RNA Pol II occupancy selectively at KDM5B target genes. We propose that KDM5B activates a self-renewal-associated gene network by repressing intragenic cryptic initiation and maintaining a H3K4me3 gradient important for productive elongation.

Proteomic analysis of the KDM5B demethylase complex reveals additional H3K36me3 and transcriptional elongation-associated proteins, including H3K36 methylation writer and reader protein NSD3. Our preliminary data demonstrate that NSD3 can mediate the recruitment of KDM5B to intragenic regions at KDM5B target genes.

We show that KDM5B can occupy and activate genes involved in cell cycle and DNA synthesis control and that forced expression of genes in the DNA synthesis machinery can partially rescue the proliferation defect after *Kdm5b* knockdown. KDM5B

is also found to interact with the DNA replication machinery, suggesting a positive feedback loop to reinforce KDM5B's role in regulating DNA replication.

Taken together, these results provide novel insights into the role of *Kdm5b* in regulating the ESC naïve pluripotency and transcriptional elongation. *Kdm5b* mRNA is also up-regulated in multiple human cancers and implicated in cancer cell proliferation and self-renewal. Thus, KDM5B may represent an attractive drug target for tumor therapy.

CHAPTER ONE

Introduction

A Brief History of Embryonic Stem Cell Research

The study of pluripotent cells dates back to the characterization of teratomas (benign) and teratocarcinomas (malignant), which are commonly found in adult gonads and are comprised of disorganized tissues from three primary germ layers: ectoderm, mesoderm and endoderm (Solter, 2006). The low rate of clinical occurrence significantly impeded the study of these tumors. In 1954, Stevens and Little found that the incidence of spontaneous testicular teratomas could reach an average of 1% in the 129 mouse strain (Stevens & Little, 1954), providing an experimental platform to study teratomas. Mice with genetic ablation of primordial germ cells fail to develop teratomas, suggesting that primordial germ cells are the origin of spontaneous teratomas (Stevens, 1967).

Intraperitoneally grafted teratocarcinomas contain poorly developed structures called embryo bodies composed of inner embryonal carcinoma cells and the outer visceral yolk sac, reminiscent of the early embryonic development. Pierce and colleagues showed that intraperitoneal injection of a single embryonal carcinoma cell could produce a teratocarcinoma (Kleinsmith & Pierce, 1964), suggesting the pluripotent stem cell nature of teratocarcinomas. Stevens showed that grafting 2-cell, pre-implantation or post-implantation embryos at extrauterine sites could produce disorganized clusters of undifferentiated embryonic cells, which in turn became teratocarcinomas (Stevens, 1968; Stevens, 1970). Stevens suggested that undifferentiated pluripotent cells, from either primordial germ cells or embryos, were the origin of teratomas or teratocarcinomas (Stevens, 1970).

In parallel with mouse models, attempts were made to establish cell lines derived from teratocarcinomas to recapitulate their biological identity *in vitro*. Martin and Evans observed that two major heterogeneous cell types coexisted in teratocarcinoma clones—

small compact cells with large nuclei and prominent nucleoli, and large spread monolayer cells. Therefore they introduced feeder cells to reliably subclone teratocarcinoma cell lines (Martin & Evans, 1974). These cell lines, termed embryonal carcinoma cells (ECCs), can proliferate indefinitely *in vitro* and reproduce teratocarcinomas upon subcutaneous injection. They can be induced to differentiate into multiple cell types mimicking embryonic development through the embryonic body formation (Martin & Evans, 1975). However, ECCs were reported to have abnormal karyotypes (Papaioannou et al, 1978), limited ability to form chimaeras and a propensity for recurrent tumorigenesis (Papaioannou et al, 1978; Rossant & McBurney, 1982), suggesting that ECCs do not recapitulate the full developmental potential of their embryonic counterparts. In 1981, Evans and Martin independently reported the success of deriving the undifferentiated cells—coined embryonic stem cells (ESCs) from the pre-implantation mouse blastocysts using ECC culture conditions (Evans & Kaufman, 1981; Martin, 1981). ESCs can self-renew *in vitro* and can produce teratomas. Importantly, ESCs are karyotypically normal and can more readily form chimaeras and contribute to germ-line transmission than ECCs (Bradley et al, 1984), suggesting these cells are truly a functional counterpart to pluripotent stem cells during embryogenesis.

The ability of mouse ESCs to contribute to germ-line chimaerism prompted the development of gene targeting via homologous recombination (Capecchi, 2005), making mouse ESCs an invaluable tool to study gene function. The study of mouse ESCs also inspired the development of a methodology that would allow the derivation of human ESCs (Thomson et al, 1998). Compared to mouse ESCs, the molecular mechanisms for establishing and maintaining human ESCs are less well understood and therefore the derivation efficiency of human ESCs still remains low (Niakan et al, 2012). In addition, human ESCs grow much slower than mouse ESCs (Singh & Dalton, 2009) and are

difficult to propagate as trypsinized single cells due to dissociation-induced apoptosis (Ohgushi et al, 2010), and display low efficiency for transgene modification (Ptaszek & Cowan, 2007). Accordingly, mouse ESCs represent an invaluable model system to study the mammalian gene function in embryogenesis, development and cellular reprogramming.

Pluripotent Stem Cell Characteristics

Self-renewing ESCs are a powerful system to model early developmental processes because they are able to give rise to all tissues and organs, a defining feature termed pluripotency. The study of pluripotency ushered in two active areas of study: derivation of ESCs from natural occurring embryogenesis (Thomson et al, 1998) and induction of pluripotency from somatic cells (Takahashi et al, 2007; Takahashi & Yamanaka, 2006). Studies on how the pluripotent state is derived, maintained and induced to differentiate have deepened our understanding of mammalian development, cellular reprogramming, and developing strategies for regenerative medicine.

Totipotency is defined as the ability of the fertilized egg to derive all cell types of an organism, including extraembryonic lineages important for the embryo development. Totipotent cells undergo rapid divisions and gradually restrict their developmental potential by entering the first cellular segregation into the inner cell mass (ICM) and the outer layer of trophectoderm at the late morula stage (Dejosez & Zwaka, 2012), accompanied by the mutually exclusive expression of two essential transcription factors—Oct4 and Cdx2, respectively (Niwa et al, 2005). Trophectoderm cells constitute the majority of the fetal part of the placenta that supports embryo growth. The ICM

continues to grow and undergoes the second segregation before implantation, forming epiblast (or primitive ectoderm) and hypoblast (or primitive endoderm), as evidenced by the segregated expression of key transcription factors Nanog and Gata6, respectively (Chazaud et al, 2006). Epiblast contributes to all the tissues and organs of the developing fetus while the primitive endoderm forms the extraembryonic endoderm layers of the visceral and parietal yolk sac. Unlike the totipotent zygote and blastomere, epiblast cells are termed pluripotent because they can give rise to all somatic and germ lineages except for the extraembryonic lineages (trophectoderm and primitive endoderm)(Dejosez & Zwaka, 2012).

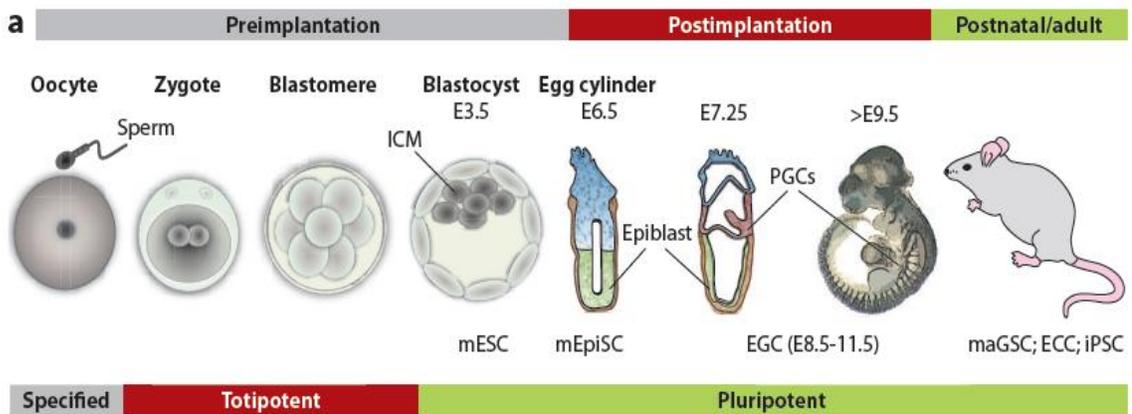


Figure 1.1: Origin of pluripotent stem cells at different stages of the mouse embryonic development (Dejosez & Zwaka, 2012). Mouse ESCs are derived from the ICM of E3.5 pre-implantation embryos while mouse EpiSCs from the epiblast of E6.5 post-implantation egg cylinder stage. Other sources of pluripotent stem cells such as mEGCs (mouse embryonic germ cells), iPSCs (induced pluripotent stem cells), ECCs (embryonic carcinoma cells) are also shown. The publisher grants reprint permission.

Mouse ESCs were originally derived from the ICM of pre-implantation blastocysts (around E3.5-E4.5 dpc) (Evans & Kaufman, 1981; Martin, 1981). Recently pluripotent

epiblast stem cells (EpiSCs) were established from post-implantation epiblasts (E5.5-E7.5 dpc) (Brons et al, 2007; Tesar et al, 2007), suggesting that multiple pluripotent states exist *in vitro*. Mouse ESCs self-renew in response to LIF and BMP4 signaling pathways (discussed in the next section), possess two active X chromosomes (female cells), readily colonize the ICM and form chimaeras (Nichols & Smith, 2009). In contrast, EpiSCs require the addition of activinA and bFGF, have an inactive X chromosome, and contribute to chimaerism and germ-line transmission with markedly reduced efficiency (Brons et al, 2007; Tesar et al, 2007). Austin Smith and colleagues proposed that two functionally distinct pluripotent states exist: 'naïve' LIF-dependent state as exemplified by mouse ESCs which readily integrate into the ICM and contribute to chimaeras and the germ line; and a 'primed' bFGF-dependent state that corresponds to EpiSCs (Nichols & Smith, 2009). Interestingly, mouse EpiSCs are very similar to human ESCs in terms of morphology, growth factor requirement, epigenetic state (X inactivation) and increased cell cycling time. It is now widely believed that human ESCs are the human correlate of EpiSCs and this has led to studies seeking to generate a *bona fide* naïve human ESC line.

Embryonic germ cells (EGCs) represent another form of pluripotent stem cells and are derived from primordial germ cells (PGCs) of the genital ridges between E10.5 and E12.5 dpc (Matsui et al, 1992; Resnick et al, 1992). They share marked similarity to mouse ESCs in terms of morphology and gene expression signatures and are able to form teratocarcinomas and chimaeric mice. This has led to the proposal that mouse ESCs may represent ICM-derived PGC-like cells (Zwaka & Thomson, 2005). Further molecular (genetic and epigenetic) characterizations of mouse ESCs, EGCs, ICM/naïve epiblast cells are needed to address the relationships between these pluripotent cells.

The naïve and primed pluripotent states are not invariant but interconvertible. ESCs can be induced to switch to EpiSCs in the presence of activinA and Fgf2 (Guo et al, 2009). EpiSCs can be converted to ESCs by forced expression of *Klf4*, *Nanog*, or *Nr5a* (Guo & Smith, 2010; Guo et al, 2009; Silva et al, 2009) or by culturing in LIF-containing ESC medium (Bao et al, 2009). Ectopic expression of constitutively active *Stat3* is required for EpiSCs to ESC conversion, strongly suggesting that LIF-Stat3 signaling is critical for the maintenance of the naïve pluripotent state (Yang et al, 2010).

The concept of a naïve pluripotent state led scientists to use reprogramming factors and small molecules to generate naïve pluripotent stem cells from rat (Liao et al, 2009) and nonpermissive mouse strains (Hanna et al, 2009). Importantly, reprogramming human somatic cells via constitutive expression of *OCT4/SOX2/KLF4/MYC/NANOG* under mouse culture conditions generated hLR5 cells (hLR5 stands for human, LIF, five reprogramming factors) that closely resembled mouse ESCs. Unlike primed human ESCs but like mouse naïve ESCs, hLR5 cells can be efficiently manipulated genetically by homologous recombination (Buecker et al, 2010). These studies suggest that the ability to generate naïve pluripotent stem cells in primates could greatly facilitate gene targeting and may prove valuable for disease modeling, developmental studies and cell-based therapy.

Signaling Pathways that Regulate Pluripotent Stem Cells

The ability of mouse ESCs to sustain a naïve pluripotent state relies on extrinsic factors that relay environmental signals to pluripotency-associated transcription factors, which in turn facilitate ESC self-renewal and antagonize differentiation.

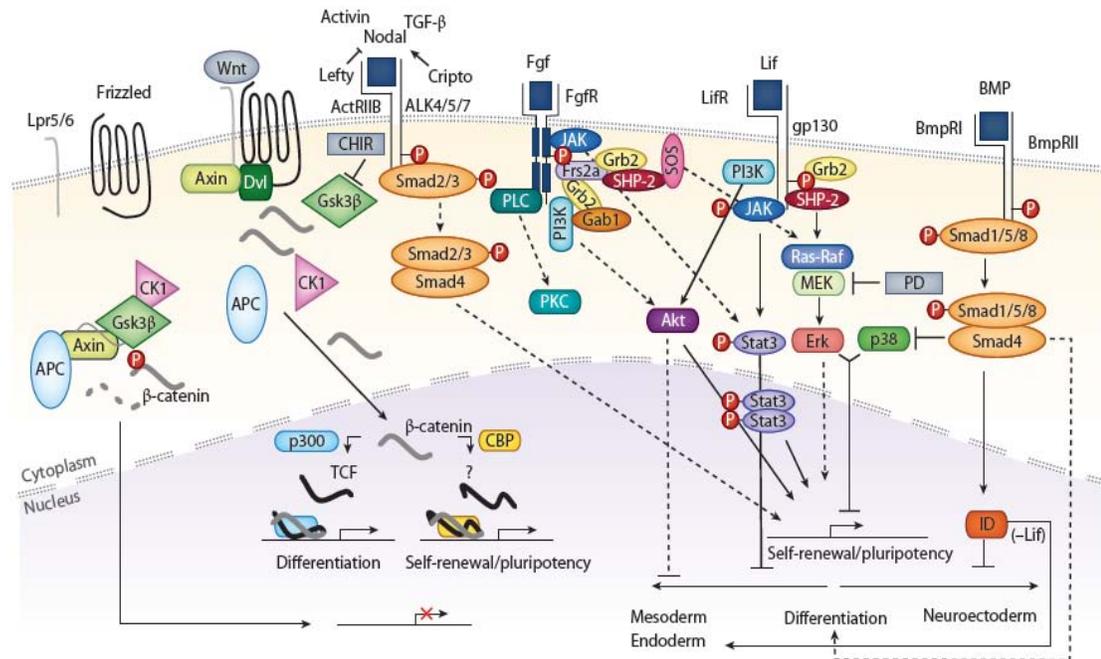


Figure 1.2: A complex signaling network regulating the pluripotent state (Dejosez & Zwaka, 2012). The major signaling pathways maintaining mouse ESCs pluripotency are LIF-Stat3 and BMPs-Smad1/5/8. The publisher grants reprint permission.

Mouse ESC derivation and self-renewal initially required proliferation-incompetent feeder layers (Evans & Kaufman, 1981; Martin, 1981). It was subsequently shown that feeder cells could secrete a soluble growth factor called leukemia inhibitory factor (LIF) that antagonized ESC differentiation (Smith et al, 1988; Williams et al, 1988). LIF functions by binding to the LIF receptor (LIFR)–gp130 heterodimer, which in turn activates the Janus kinase (Jak) and triggers Stat3 translocation into the nucleus where it regulates genes involved in ESC pluripotency (Niwa et al, 1998). Moreover, gp130-/- lack epiblast cells, indicating an essential role for the LIF-LIFR/gp130 signaling pathway in specifying pluripotent stem cells in the mouse embryo (Nichols et al, 2001).

LIF also activates the phosphatidylinositol 3-OH kinase (PI3K)-Akt pathway and the mitogen-activated protein kinase (Mapk) pathway (Niwa et al, 2009). The PI3K-Akt pathway is required for the ESC maintenance because treatment of the PI3K-specific inhibitor reduces the ability of LIF to promote ESC self-renewal (Paling et al, 2004). In contrast, the Mapk pathway negatively regulates ESC self-renewal *in vitro* (Kunath et al, 2007) and is required for specifying the primitive endoderm in mouse blastocysts (Chazaud et al, 2006). Thus, LIF activates parallel signaling pathways to maintain ESC pluripotency. Niwa and colleagues provide further molecular insights into how LIF-activated signaling pathways regulate pluripotency-associated transcriptional pathways (Niwa et al, 2009). LIF-activated Jak–Stat3 and PI3K-Akt pathways activate Klf4 and Tbx3, which in turn activate Sox2 and Nanog. Forced expression of Klf4 or Tbx3 can confer LIF-independent maintenance of pluripotency. However, overexpression of Nanog supports LIF-independent mouse ESC self-renewal even in the absence of Klf4 and Tbx3 activity, suggesting that Klf4 and Tbx3 function to mediate the LIF signaling to the core pluripotency factor Nanog (Niwa et al, 2009).

LIF is insufficient for ESC maintenance in the absence of serum. Ying and colleagues show that bone morphogenetic protein 4 (BMP4) is the component in serum that is required for ESC self-renewal (Ying et al, 2003). BMPs belong to the TGF- β superfamily, bind to heterodimerized type I and type II serine/threonine kinase receptors, which in turn activate the downstream Smad transcription factors. Activated Smad1/5/8 translocate into the nucleus to stimulate the expression of inhibitor of differentiation (*Id*) genes to suppress the neural lineage commitment (Ying et al, 2003) and antagonize the detrimental effect of the Mapk signaling on ESC self-renewal (Qi et al, 2004).

Genetic Network of Core Pluripotency Transcription Factors

The first identified, and perhaps most essential transcription factor to ESC pluripotency is Oct4 (octamer-binding transcription factor 4, encoded by *Pou5f1*). *Oct4* is highly expressed in the totipotent embryos and the pluripotent ICM, ESCs, ECCs and germ cells (Palmieri et al, 1994; Yeom et al, 1996). Genetic deletion of *Oct4* disrupts the pluripotent ICM formation and causes trophectoderm differentiation (Nichols et al, 1998), suggesting that *Oct4* is essential for specifying pluripotent founder cells during embryogenesis. In mouse ESCs, a narrow window of Oct4 protein expression is critical. Even a two-fold increase in Oct4 protein level differentiates ESCs into mesoderm and primitive endoderm lineages while depletion of *Oct4* leads to dedifferentiation into the trophectoderm (Niwa et al, 2000). Oct4 is also indispensable for epigenetic reprogramming (Takahashi et al, 2007; Takahashi & Yamanaka, 2006). These studies suggest that Oct4 is a master regulator of the ESC pluripotency. Oct4 associates with a large interactome network in ESCs composed of key pluripotency transcription factors, chromatin modification complexes, basal transcription machinery components and signaling pathway molecules (Ding et al, 2011; Liang et al, 2008; Pardo et al, 2010; van den Berg et al, 2010). Acute depletion of *Oct4* decreases the recruitment of pluripotency transcription factors and chromatin modifying enzymes at tested Oct4 binding loci (Chen et al, 2008; van den Berg et al, 2010), suggesting Oct4 as the core cis-acting transcription factor orchestrating the pluripotency-specific multiprotein complex to govern the ESC pluripotent state.

Oct4 can heterodimerize with another core pluripotency factor Sox2 (sex determining region Y-box 2), which is enriched in the ICM, epiblast and extraembryonic ectoderm (Avilion et al, 2003). Genetic deletion of *Sox2* results in peri-implantation

lethality of developing mouse embryos and defects of the epiblast formation (Avilion et al, 2003). Genetic targeting of *Sox2* causes ESCs to differentiate into trophectoderm and decreases *Oct4* expression. On the other hand, forced expression of *Oct4* can rescue the *Sox2* depletion phenotype. Thus, *Sox2* may stabilize the ESC pluripotent state by sustaining the robust expression of *Oct4* (Masui et al, 2007).

Niwa and colleagues found that forced expression of *Nanog* could support LIF-independent ESC self-renewal even in the absence of pluripotency transcription factor *Klf4* or *Tbx3*, suggesting that *Nanog* is the core pluripotency factor mediating the LIF pathway (Niwa et al, 2009). Distinct from *Oct4* or *Sox2*, ectopic expression of *Nanog* can even bypass the requirement of both BMP4/serum and LIF to maintain ESCs in an undifferentiated state (Ying et al, 2003). *Nanog* null embryos display peri-implantation lethality and cannot mature into epiblasts (Mitsui et al, 2003). Depletion of *Nanog* in ESCs triggers morphological differentiation and increases the expression of multiple lineage marker genes (Ivanova et al, 2006). Surprisingly, *Nanog* null ESCs can self-renew *in vitro*, suggesting that *Nanog* is dispensable for the ESC maintenance (Chambers et al, 2007). Nevertheless, *Nanog* null ESCs are predisposed to differentiate and show a PGC maturation defect (Chambers et al, 2007). In transcription factor-induced somatic reprogramming, *Nanog* is initially dispensable but is essential for final entry into a naïve pluripotent state. Thus, *Nanog* is essential for the acquisition of both embryonic and induced pluripotency (Silva et al, 2009) and is believed to specify and stabilize the naïve pluripotent state (Theunissen & Silva, 2011).

Genomic approaches using chromatin immunoprecipitation (ChIP) followed by DNA microarray hybridization or high-throughput sequencing (ChIP-Seq) were employed to gain molecular insights into how these core pluripotency factors maintain ESC identity

(Boyer et al, 2005; Loh et al, 2006). Several key insights emerged from these studies (Jaenisch & Young, 2008). First, Oct4, Sox2 and Nanog stabilize the ESC pluripotent state by forming an autoregulatory circuitry and bind their own promoters to reinforce their expression. Second, Oct4, Sox2, and Nanog share large overlapping genomic targets, suggesting that these transcription factors regulate a large set of common target genes critical for ESC maintenance. In fact, many of these pluripotency transcription factors constitute a large multiprotein complex network that coordinately regulates ESC pluripotency (Liang et al, 2008; Wang et al, 2006). Third, these master pluripotency regulators not only bind to active genes promoting ESC self-renewal, but also bind to repressed developmental regulators.

Interaction of core pluripotency factors with p300 and the mediator complex is believed to form an “enhanceosome” that activates transcription initiation (Chen et al, 2008; Kagey et al, 2010; Young, 2011). This may explain why 70% of ESC genes engage normal transcriptional initiation, but undergo abortive transcriptional elongation lacking elongating RNA Pol II and H3K36me3 (Guenther et al, 2007; Min et al, 2011). c-Myc is important for both ESC maintenance and induced reprogramming (Cartwright et al, 2005; Sridharan et al, 2009; Takahashi et al, 2007; Takahashi & Yamanaka, 2006; Varlakhanova et al, 2010). c-Myc was shown to recruit P-TEFb to release paused Pol II for productive elongation (Rahl et al, 2010). These studies suggest that pluripotency transcription factors utilize different strategies to coordinate a transcription program important for the establishment and maintenance of ESC identity.

Epigenetic Mechanisms Governing Pluripotent Stem Cells

The specification of functionally distinct cell types carrying identical genomes is essential for the development of multicellular organisms. Increasing evidence documents epigenetic mechanisms as key choreographers in instructing and maintaining cell-specific gene expression patterns (Margueron & Reinberg, 2010).

Pluripotent ESCs have a characteristic open chromatin structure that is enriched in H3K9ac. Following ESC differentiation, chromatin forms noticeable condensed blocks of heterochromatin enriched in H3K9me2 (Gaspar-Maia et al, 2010; Niwa, 2007b). A global hypertranscription phenotype is also observed in ESCs (Efroni et al, 2008). In addition, pluripotent ESCs, but not lineage committed cells or terminally differentiated cells, display hyperdynamic binding properties of key histone (H1, H2B, H3) and non-histone (HP1) proteins to chromatin (Meshorer et al, 2006).

The core pluripotency factor circuitry (Oct4/Sox2/Nanog) regulates multiple chromatin modifying enzymes required for maintaining the ESC pluripotency, including H3K9 histone lysine demethylase (KDM) Jmjd2C and Jmjd1a (Loh et al, 2007), H3K4 histone methyltransferases (HMT) complex core subunit Wdr5 (Ang et al, 2011), DNA hydroxymethyltransferase Tet1 (Freudenberg et al, 2012; Ito et al, 2010) and others. Oct4 and Nanog can also physically interact with multiple chromatin modification enzymes and chromatin remodeling complexes to actively organize the ESC epigenome favorably for pluripotency-associated transcription, DNA replication and repair (Ang et al, 2011; Ding et al, 2011; Liang et al, 2008). This suggests that core pluripotency-associated transcriptional pathways may help facilitate a naïve epigenetic state.

Polycomb group proteins (PcG) deposit H3K27me3 and are believed to facilitate gene repression that contributes to developmental specification (Margueron & Reinberg, 2011). Interestingly, many of the developmental regulatory genes are marked by both repressive H3K27me3 and active H3K4me3 at their promoters, a chromatin signature termed the bivalent domain (Bernstein et al, 2006). The bivalent modifications may contribute to pluripotency by enhancing the plasticity to switch cell fates in response to developmental signals to warrant robust differentiation.

Reprogramming of somatic cells to the pluripotent state requires a cocktail of factors: Oct4, Sox2, Klf4, c-Myc for mouse and OCT4, SOX2, NANOG, and LIN28 for human (Takahashi et al, 2007; Takahashi & Yamanaka, 2006; Yu et al, 2007). Reprogramming is accompanied by global changes in the somatic epigenome, including inactive X chromosome reactivation, DNA demethylation at pluripotency gene promoters and resetting of the bivalent modifications (Hochedlinger & Plath, 2009). It is not known whether these epigenetic changes are simply consequences of cellular reprogramming or play active roles in the acquisition of pluripotency. A loss-of-function survey toward chromatin modifying enzymes in human iPSCs formation demonstrated that inhibition of PcG components (EZH2, SUZ12, EED, BMI1, RING1) or H3K9 HMTs (SETDB1 and EHMT1) significantly decreased the reprogramming efficiency whereas knockdown of H3K79 HMT DOT1L or another H3K9 HMT SUV39H1 increased the reprogramming efficiency mediated by the human OCT4/SOX2/KLF4/ MYC combination (Onder et al, 2012). DOT1L catalyzes formation of H3K79me2 which is associated with the active transcriptional elongation of a fibroblast-specific gene expression program. Inhibition of DOT1L suppressed the H3K79me2-marked active chromatin and increased mesenchymal-to-epithelial transition. This study demonstrated the causal roles of chromatin modifying enzymes as barriers or facilitators of cellular reprogramming by

acting on the local chromatin architecture (Onder et al, 2012). In addition, a proteomics approach identified ATP-dependent BAF (Brg1 associated factors) chromatin remodeling complex which could greatly enhance the reprogramming efficiency presumably through rearranging nucleosomes and augmenting Oct4 binding at target gene promoters (Singhal et al, 2010).

Taken together, these studies suggest that epigenetic mechanisms are intrinsically hard-wired into the acquisition and maintenance of the pluripotency program and understanding the molecular epigenetic mechanisms instating pluripotency will have invaluable impacts on cellular reprogramming research and regenerative medicine.

Nanog Connects Genetic and Epigenetic Programs for ESC Identity

Unlike Oct4 or Sox2 (Masui et al, 2007; Niwa et al, 2000), forced *Nanog* expression can support mouse ESC self-renewal independent of the LIF signaling (Chambers et al, 2003; Mitsui et al, 2003). Moreover, two LIF-activated parallel signaling pathways (Jak-Stat3-Klf4 and PI3K-Akt-Tbx3) converge on Nanog (Niwa et al, 2009). Interestingly, Nanog shows heterogeneous expression in mouse ESCs (Chambers et al, 2007; Kalmar et al, 2009) and Nanog-GFP ESCs can be sorted into high Nanog (HN) and low Nanog (LN) subpopulations. The percentage of the HN portion increases in the naïve ESC state (Silva et al, 2009) whereas LN subpopulations are unstable and prone to differentiate, reminiscent of a primed pluripotent state (Chambers et al, 2007; Kalmar et al, 2009). In addition, Nanog overexpression facilitates EpiSCs to ESCs conversion (Silva et al, 2009). These observations suggest that the stoichiometry of Nanog can

specify different pluripotent states and that high *Nanog* expression specifically inaugurates naïve pluripotency (Theunissen & Silva, 2011).

In cell fusion-mediated reprogramming between ESCs and neural stem cells (NSCs), elevated levels of *Nanog* markedly increased the number of hybrids with ESC-like colony properties (Silva et al, 2006). Depletion of *Nanog* largely abolished the ability of ESCs to reprogram NSCs during cellular fusion (Silva et al, 2009). Therefore, *Nanog* is required and sufficient for the NSC epigenome to be reset to a pluripotent state. During reprogramming of NSCs, extensive epigenetic remodelling occurs, including reactivation of inactive X chromosomes, resetting of bivalent modifications, loss of DNA methylation at core pluripotency factor promoters (Hochedlinger & Plath, 2009). This led to the proposal that *Nanog* choreographs an epigenetic program that wipes out the somatic epigenome and dictates pluripotency. *Nanog* is also highly expressed in primordial germ cells (PGCs) (Yamaguchi et al, 2005) and is required for PGCs maturation in the germ lineage (Chambers et al, 2007). Profound epigenetic changes occur when PGCs colonize the genital ridge including genomic imprinting erasure, global DNA demethylation, and re-activation of the inactive X chromosomes in female cells. Ian Chambers posited that *Nanog* is also essential for the induction of a pluripotent state in germ cells (Chambers et al, 2007).

Despite the general consensus that *Nanog* is required for the naïve pluripotent state (Nichols & Smith, 2009), the molecular mechanisms by which *Nanog* initiates and regulates naïve ESCs identity are elusive. Although the genome-wide snapshot of *Nanog* occupancy has been published in ESCs (Boyer et al, 2005; Loh et al, 2006; Marson et al, 2008), the downstream effector genes, particularly epigenetic modulator

genes that mediate Nanog functionality in achieving naïve pluripotent state have not been experimentally tested.

In the current study, we aim to bridge the Nanog mechanisms and downstream epigenetic mechanisms by asking a central question: ***what are the downstream epigenetic mechanisms of Nanog in orchestrating naïve pluripotency?*** We take advantage of functional genomics approaches developed and extensively used in the Impey lab. We generated high quality Nanog ChIP-Seq libraries and Nanog siRNA knockdown (siNanog vs siControl) RNA-Seq libraries from multiple biological replicates using Illumina next-generation genome DNA sequencing technique (Metzker, 2009).

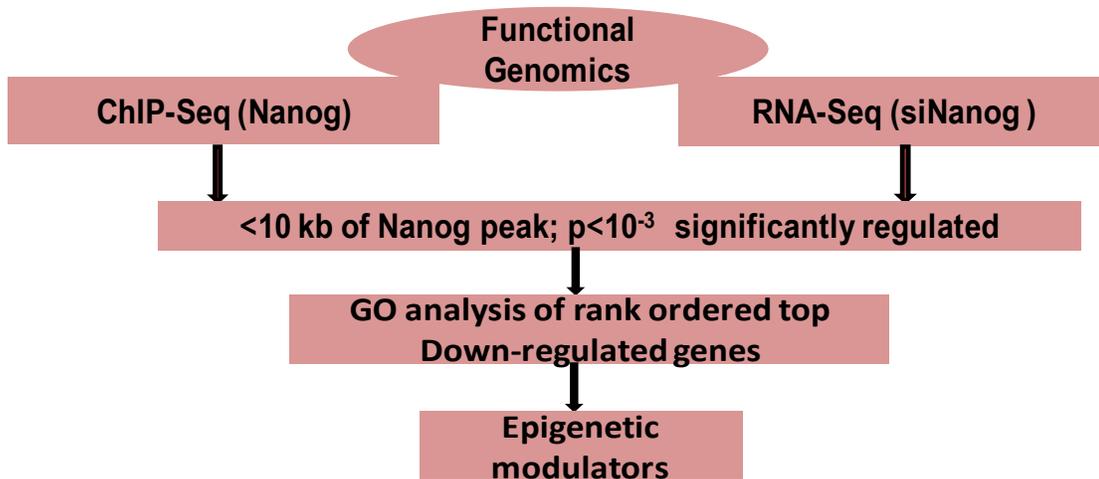


Figure 1.3: Flow chart of functional genomics approaches to interrogate Nanog-regulated epigenetic mechanisms towards ESC naïve pluripotency. Nanog ChIP-Seq and siNanog RNA-Seq libraries were generated using Illumina Solexa next-generation genomic sequencer (GAII). Intersection of the two libraries generated a list of genes that are bound within 10kb of Nanog binding peaks and significantly regulated ($p < 10^{-3}$) by Nanog knockdown. Gene ontology analysis using DAVID (<http://david.abcc.ncifcrf.gov/>) identified genes with chromatin modulation functions. These chromatin regulatory genes rank-ordered by transcript fold change following Nanog knock down are shown in Table 1.1.

By intersecting the Nanog ChIP-Seq and siNanog RNA-Seq libraries, we select genes that are significantly bound by Nanog and positively regulated by Nanog. Because we aim to understand the Nanog downstream epigenetic mechanisms, we utilize Gene Ontology analysis (DAVID) of RefSeq genes and custom annotation to identify significantly enriched genes that encode chromatin modulators, including histone modification enzymes (writers, erasers, effectors), DNA methyltransferases or hydroxymethyltransferases and chromatin remodeling complexes. We propose that chromatin regulatory genes bound and positively regulated by Nanog may mediate Nanog functionality in shaping and maintaining an epigenetic landscape favorable to naïve pluripotency.

We focus on a list of 26 putative Nanog target genes encoding chromatin modulators, which can be broadly categorized into three groups.

Group one--known chromatin modulator genes involved in ESC maintenance (6 out of 26), including *Rcor2* (Yang et al, 2011), *Wdr5* (Ang et al, 2011), *Tet1* (Freudenberg et al, 2012; Ito et al, 2010), *Arid1b* (Yan et al, 2008), *Phc1* (Walker et al, 2007) and *Kdm3a* (Loh et al, 2007). For example, *Rcor2* is part of LSD1 complex and its overexpression enhances reprogramming by replacing Sox2 (Yang et al, 2011). Both H3K4 HMT complex component *Wdr5* (Ang et al, 2011) and DNA hydroxymethyltransferase *Tet1* (Freudenberg et al, 2012; Ito et al, 2010) were shown to be epigenetic components essential for ESC maintenance.

Group two—known chromatin modulator genes which are not involved in ESC maintenance (6 out of 26), including *Dnmt3a* (Tsumura et al, 2006), *Msh6* (Edelmann et al, 1997), *Mbd3* (Kaji et al, 2006; Zhu et al, 2009), *Jarid2* (Landeira et al, 2010), *Ehmt2*

(Tachibana et al, 2002), *Chd7* (Schnetz et al, 2010). Instead, *Dnmt3a*, *Mbd3* and *Jarid2* are required for ESC differentiation. Although DNA helicase *Chd7* knockdown does not produce noticeable ESC self-renewal phenotype, *Chd7* was shown to co-occupy the Oct/Sox2/Nanog enhanceosome in ESCs (Schnetz et al, 2010).

	Fold Change	Q-value	Gene Symbol	Gene Name	Classification
1	-5.57	2.04E-015	<i>Myst4</i>	MYST histone acetyltransferase	chromatin writer
2	-4.68	2.04E-015	<i>Rcor2</i>	REST corepressor 2	chromatin eraser
3	-3.27	6.42E-005	<i>Kdm4a</i>	jumonji domain containing 2A	chromatin eraser
4	-3.07	2.04E-015	<i>Setd5</i>	SET domain containing 5	chromatin writer
5	-2.99	2.04E-015	<i>Dnmt3a</i>	DNA methyltransferase 3A	chromatin writer
6	-2.84	4.01E-015	<i>Phf15</i>	PHD finger protein 15	chromatin reader
7	-2.71	2.04E-015	<i>Kdm5b</i>	Lysine demethylase 5B	chromatin eraser
8	-2.69	2.04E-015	<i>Tet2</i>	Ten-eleven translocation 2	chromatin writer
9	-2.66	2.04E-015	<i>Wdr62</i>	WD repeat domain containing 62	chromatin writer?
10	-2.56	2.04E-013	<i>Cxxc5</i>	CXXC finger 5	chromatin reader
11	-2.23	2.04E-015	<i>Wdr5</i>	WD repeat domain containing 5	chromatin writer
12	-2.13	2.04E-015	<i>Phf20</i>	PHD finger protein 20	chromatin reader
13	-2.11	1.17E-014	<i>Arid5b</i>	AT rich interactive domain 5B	chromatin remodeler
14	-2.10	2.04E-015	<i>Kdm3b</i>	Lysine demethylase 3B	chromatin eraser
15	-2.08	2.04E-015	<i>Tet1</i>	Ten-eleven translocation 1	chromatin writer
16	-1.96	2.04E-015	<i>Msh6</i>	mutS homolog 6	chromatin reader
17	-1.90	2.04E-015	<i>Mbd3</i>	methyl-CpG binding domain protein 3	chromatin reader
18	-1.89	2.04E-015	<i>Jarid2</i>	Jumonji protein 2	chromatin effector
19	-1.88	1.14E-005	<i>Arid1b</i>	AT rich interactive domain 1B	chromatin remodeler
20	-1.87	2.04E-015	<i>Phc1</i>	Polyhomeotic-like 1 isoform a	chromatin reader
21	-1.84	2.04E-015	<i>Kdm3a</i>	Lysine demethylase 3A	chromatin eraser
22	-1.68	2.44E-004	<i>Hdac5</i>	Histone deacetylase 5	chromatin eraser
23	-1.63	4.05E-013	<i>Ehmt2</i>	Euchromatic histone lysine N-methyltransferase 2	chromatin writer
24	-1.63	2.04E-015	<i>Wdr82</i>	WD repeat domain containing 82	chromatin writer
25	-1.47	8.81E-006	<i>Chd7</i>	Chromodomain helicase DNA binding protein 7	chromatin reader
26	-1.40	1.13E-002	<i>Kdm2b</i>	Lysine demethylase 2B	chromatin eraser

Table 1.1 List of Nanog-regulated chromatin modulator genes. The twenty six Nanog bound and positively regulated chromatin regulator genes in ESCs rank ordered by fold change of gene expression after Nanog knockdown were inferred from our Nanog ChIP-Seq and siNanog RNA-Seq libraries. Red color marks genes that have been studied which are required for ESC self-renewal and pluripotency (Ang et al, 2011; Freudenberg et al, 2012; Ito et al, 2010; Loh et al, 2007; Walker et al, 2007; Yan et al, 2008; Yang et al, 2011). Black color marks genes which are not required for ESC maintenance but may involve in ESC differentiation (Edelmann et al, 1997; Kaji et al, 2006; Landeira et al, 2010; Schnetz et al, 2010; Tachibana et al, 2002; Tsumura et al,

2006). Green colors mark genes whose functions in ESCs have not been studied or poorly understood.

Group three—chromatin modulator genes whose biological functions have not been studied or are poorly understood in the context of ESC self-renewal and pluripotency (14 out of 26), including *Myst4*, *Kdm4a*, *Setd5*, *Phf15*, *Kdm5b*, *Tet2*, *Wdr62*, *Cxxc5*, *Phf20*, *Arid5b*, *Kdm3b*, *Hdac5*, *Wdr82*, *Kdm2b*.

To assess their functional contributions to ESC identity, I systematically studied their individual requirement for ESC maintenance by RNAi. I also designed a functional expression assay that assessed whether ectopic expressing each of them can support ESC pluripotency in the absence of LIF signaling. Both assays converged on a common set of novel chromatin modulator genes—*Myst4*, *Phf15*, *Kdm5b* and *Wdr82* which are important to maintain ESCs in an undifferentiated, pluripotent state (a detailed description of the assays and results will be discussed in Chapter Three).

Initially I focused primarily on *Kdm5b* because it is one of the top Nanog binding targets from two independent Nanog ChIP-Seq libraries. In addition, *Kdm5b* expression is markedly upregulated in multiple human cancers and is implicated in cancer cell proliferation, cell cycle control or self-renewal (Hayami et al, 2010; Roesch et al, 2010; Xiang et al, 2007; Yamane et al, 2007). I will describe the chromatin mechanisms of *Kdm5b* in ESCs in Chapter Two (paper published on EMBO J). In Chapter Three, I will present evidence to show the critical role of *Kdm5b* in regulating ESC pluripotency.

KDM5 Family of Histone Demethylases

1. Mechanisms of Histone demethylation

Unlike histone deacetylation catalyzed by HDACs that involves simple hydrolysis of the amide bond, histone lysine or arginine methylation (N-C bond) was long held to be stable and irreversible. Until 2004, the first *bona fide* histone demethylase lysine specific demethylase 1 (LSD1/KDM1) was discovered (Shi et al, 2004). LSD1 utilizes a specified amine oxidation mechanism involving reduced FAD (flavin adenine dinucleotide) as cofactor to demethylate mono and di-methyl lysine 4 or lysine 9 and produces formaldehyde as a by-product (Metzger et al, 2005; Shi et al, 2004).

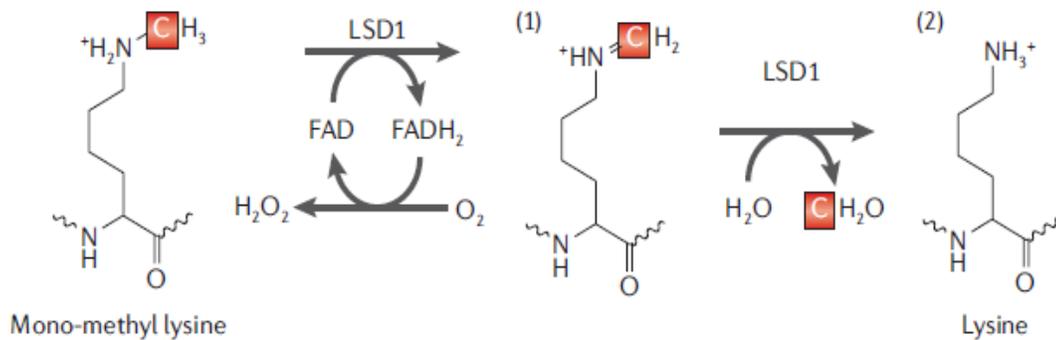


Figure 1.4: Biochemical mechanisms of LSD1-mediated histone demethylation.

Figure was adopted from (Klose et al, 2006). LSD1 is proposed to mediate demethylation of mono/di-methylated lysine residue via amine oxidative mechanism that utilizes FAD as a cofactor. The publisher grants reprint permission.

In 2005, Dr. Robin Allshire and colleagues proposed an intriguing hypothesis that, Jumonji C (JmjC) domain-containing, Fe(II) and α -ketoglutarate (α KG) dependent dioxygenases can reverse lysine methylation in a mechanism similar to the bacteria AlkB DNA repair enzyme (Trewick et al, 2005). Milestone studies from Dr. Yi Zhang's lab

provided the first experimental evidence that JHDM1A (Jumonji domain-containing histone demethylase 1, FBXL11/KDM2A) possesses H3K36me2/me1 demethylase activity by designing an elegant *in vitro* demethylation assay detecting the end product formaldehyde (Tsakada et al, 2006) (Figure 1.5). Jumonji domain-containing proteins constitute a large cupin superfamily that catalyze the demethylation of alkylated DNA in bacteria (Falnes et al, 2002; Trewick et al, 2002), conversion of 5-methylcytosine to 5-hydroxymethylcytosine in the mammalian genome (Tahiliani et al, 2009), and hydroxylation of non-histone proteins (Webby et al, 2009).

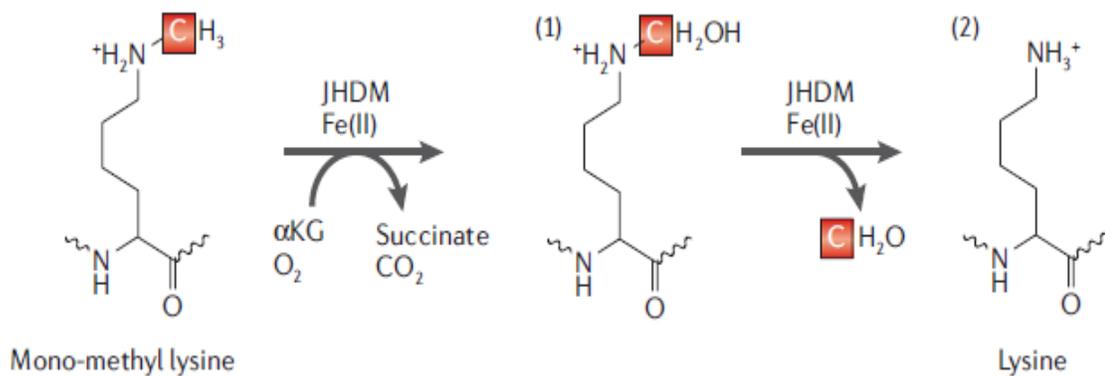


Figure 1.5: Biochemical mechanisms of JmjC domain-containing histone lysine demethylases (KDMs). Figure was adopted from (Klose et al, 2006). JHDM (JmjC domain-containing histone demethylase) or KDM can mediate demethylation of mono/di/tri-methylated lysine (or arginine) residue via an oxidative mechanism that requires Fe(II) and α KG as cofactors. For simplicity, only the demethylation of mono-methylated lysine substrate is shown. The publisher grants reprint permission.

2. KDM5 family member domain architecture, phylogenetic relationship and structural analysis

Thereafter, multiple JmjC domain-containing proteins have been discovered in mammalian cells that possess demethylase activity towards different lysine residues with different methylation states (tri-, di- or mono-methylation) on histone tails (Klose et al, 2006). These proteins are collectively named KDMs (lysine demethylases), and classified into six major subfamilies based on their catalytic activity according to the new nomenclature (Allis et al, 2007). KDM5/JARID1 family members possess specific histone 3 lysine 4 tri- and di-methylation (H3K4me2/me3) demethylase activity and conserved protein domains including JmjN, Arid, JmjC, PHD and C5HC2 zinc finger demonstrated in Figure 1.6 (Christensen et al, 2007; Iwase et al, 2007; Klose et al, 2007; Lee et al, 2007a; Yamane et al, 2007).

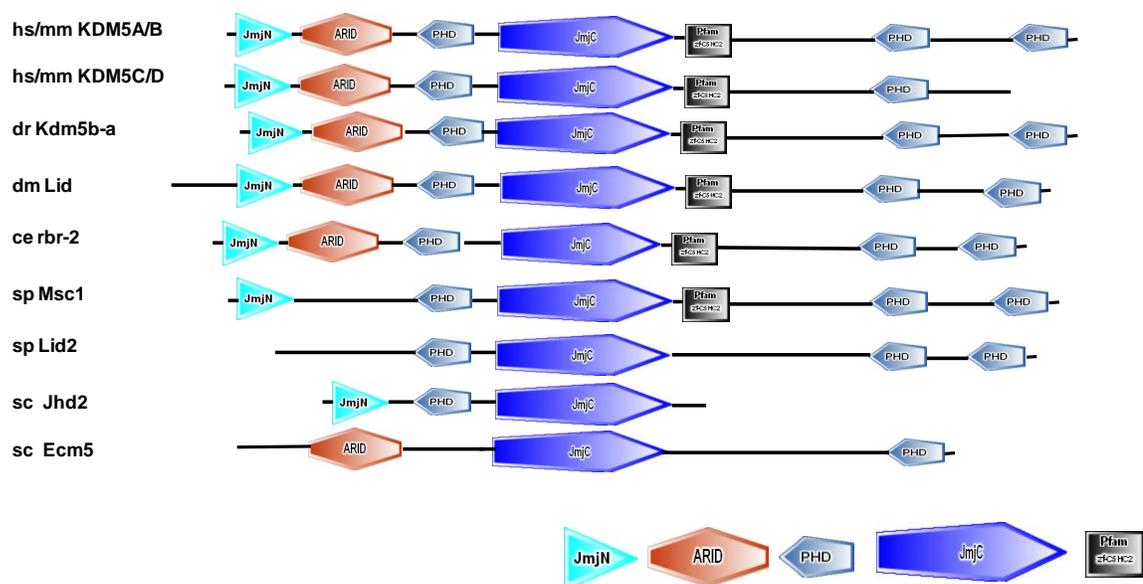


Figure 1.6: Domain architecture comparison of KDM5/JARID1 histone demethylase proteins. The domain architecture among KDM5/JARID1 family members in different model organisms is conserved and contains ordered arrangement of Jumonji N (JmjN), ARID, Jumonji C (JmjC), plant homeodomain (PHD) and C5HC2-Zinc Finger domains. Domain architecture were retrieved from PFAM database and analyzed by SMART program (Simple Modular Architecture Research Tool) (Letunic et al, 2011).

Based on KDM5 family member sequence homology and domain architecture similarity, I inferred a phylogenetic history tree of KDM5 demethylase family members (Figure 1.7) based on their amino acid sequence information. As predicted, KDM5/JARID1 family members show clear evolutionary segregation with JARID2 family members. And to our interest, among the four mammalian KDM5 family members, Kdm5b seems to have a separate branching with the rest of KDM5 homologues.

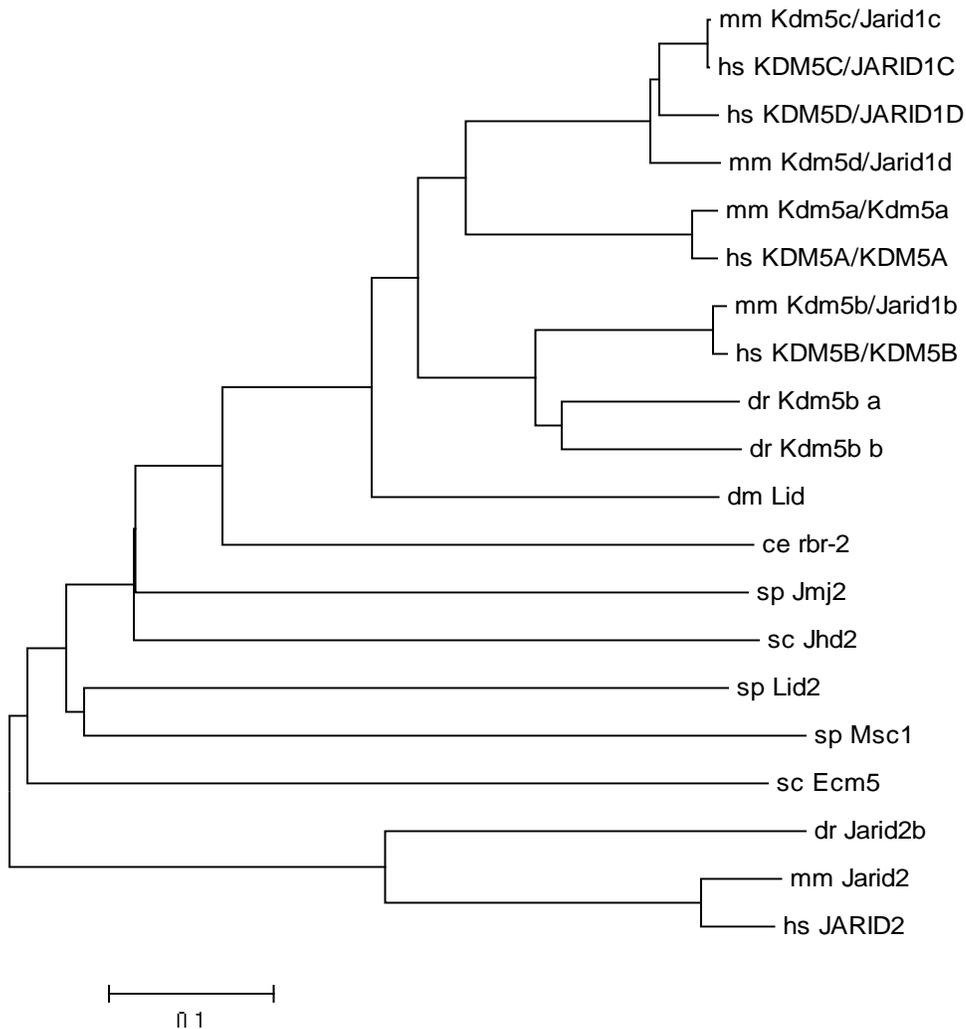


Figure 1.7: De novo generation of phylogenetic relationship among KDM5/JARID1 family proteins from model organisms. The protein sequences of KDM5/JARID1 family members were extracted from PFAM protein family database (Punta et al, 2011) in model organisms human (*Homo sapiens*, hs), mouse (*Mus musculus*, mm), zebrafish (*Danio rerio*, dr), fly (*Drosophila melanogaster*, dm), worm (*Caenorhabditis elegans*, ce), budding yeast (*Saccharomyces cerevisiae*, sc) and fission yeast (*Schizosaccharomyces pombe*, sp). A multiple sequence alignment was carried out using the ClustalW program (Chenna et al, 2003) followed by phylogenetic-tree analysis using the Molecular Evolutionary Genetic Analysis (MEGA5) (Tamura et al, 2011) and the Neighbor-Joining method with the sum of branch length (SBL) = 4.44628679. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and were in the units of the number of amino acid differences per site (Tamura et al, 2011).

The crystal structure of the JMJD2C/KDM4 family members has been resolved, providing insight into the molecular mechanisms of histone demethylation by JmjC family members (Chen et al, 2006). Eight β -sheets within the catalytic core form a jellyroll-like structure resembling the cupin metalloenzyme superfamily (Klose et al, 2006). In the native structure of the JmjC domain of JMJD2A/KDM4A, an iron atom --Fe(II) is chelated by three highly conserved residues-- His188, Glu190, and His276, consistent with the biochemical studies that a His-X-ASP/GLU-Xn-His signature motif is essential for coordinating the Fe(II) cofactor and mutation of which largely abolishes the histone demethylase activity (Tsukada et al, 2006; Whetstone et al, 2006). Structural studies demonstrate that the other cofactor α -KG is associated with the Fe(II) ion through the C-1 carboxylate and C-2 ketone groups, and is further stabilized by three hydrogen bonds formed between α -KG and the side chains of Tyr132, Asn198, and Lys206 of JMJD2A (Chen et al, 2006). The crystal structure of the highly conserved JmjC domain of

JMJD2A and the highly homologous amino acids residues within KDM5/JARID1 family members are depicted below in Figure 1.8.

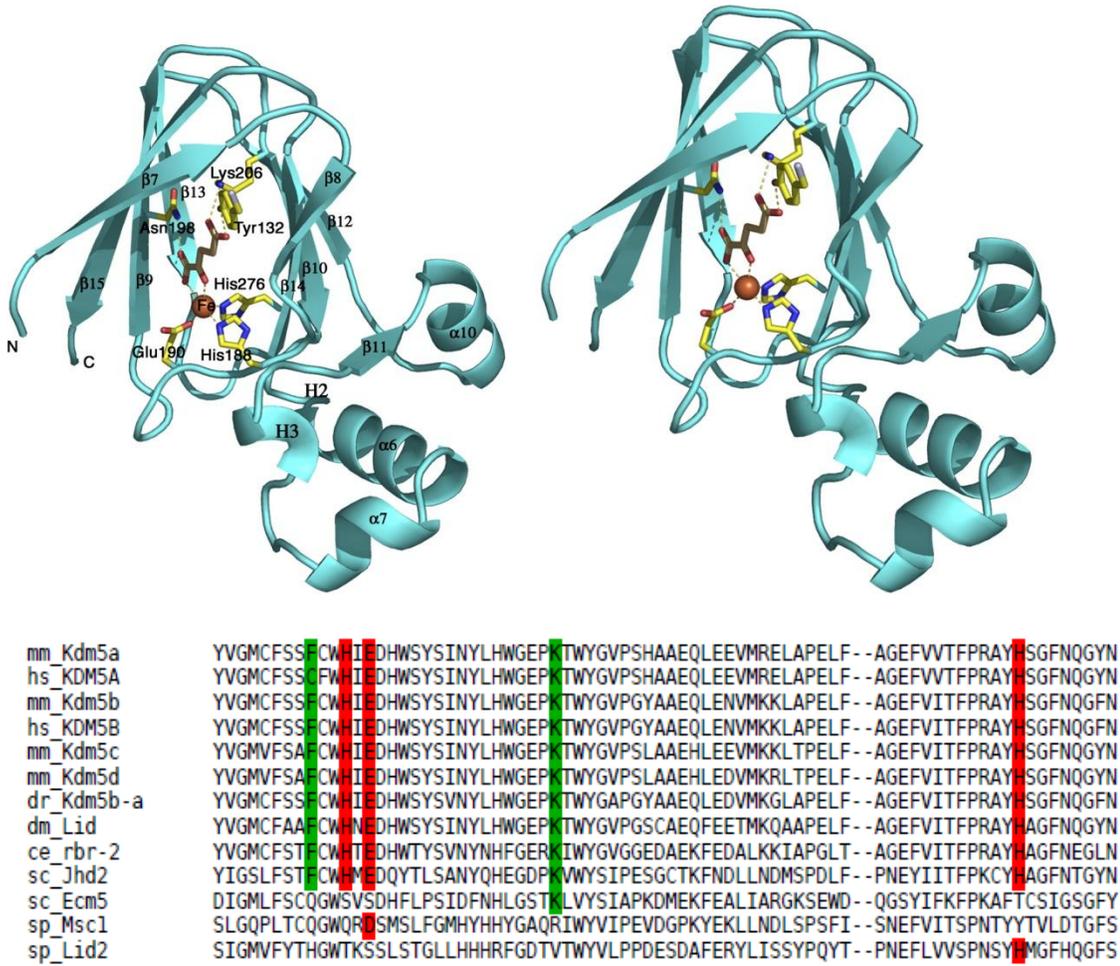


Figure 1.8: Structural analysis of catalytic JmjC domain. **A** The crystal structure of JmjC domain from JMJD2A. Residues His188, Glu190, and His276 are involved in chelating the Fe(II) atom (labeled in red ball). Tyr132, Asn198, and Lys206 contribute to stabilizing the α -KG cofactor (figure was adopted from (Chen et al, 2006)). **B** Multiple sequence alignment using ClustaW (Chenna et al, 2003) demonstrates highly conserved residues among KDM5/JARID1 family members contributing to the binding of Fe(II) (in red) and α -KG (in green). The publisher grants reprint permission.

In addition to the catalytic JmjC domain, the JmjN, Arid, N-terminal PHD (PHD1) and the C5HC2 zinc finger domains are all required for the demethylase activity (Li et al, 2010; Yamane et al, 2007). PHD1 can bind methylated H3K9 (Iwase et al, 2007; Li et al, 2010; Wang et al, 2009a) or unmethylated H3K4 *in vitro* (Li et al, 2010). Mouse Kdm5a, Kdm5b and their Drosophila homolog Lid all have two consecutive C-terminal PHD domains (PHD2 and PHD3) while Kdm5c and Kdm5d have only one (PHD2). PHD3 can specifically recognize and bind H3K4me_{2/3} but is dispensable for enzymatic activity (Li et al, 2010; Li et al, 2011; Wang et al, 2009a; Yamane et al, 2007). The Lid-PHD3 domain was proposed to recognize H3K4me_{2/3} and in turn recruit dMyc to E-box regions in the chromatin context to activate gene expression (Li et al, 2010). Therefore, it is important to study the domain structures and protein-protein interactions of the KDM5 family members to fully understand their biological functions.

3. The function of KDM5 family members in model organisms

The KDM5 fission yeast homologue Lid2 has been characterized as a specific H3K4me₃/me₂ demethylase and appears to have dual functions marking different chromatin domains (Li et al, 2008). Lid2 interacts with H3K9 HMT Clr4 and the RNAi pathway complex Dos1/Clr8-Rik1 to coordinately maintain heterochromatin silencing. JmjC disruption results in severe heterochromatin defects and depletion of the RNAi pathway. On the other hand, Lid2 can recruit H3K4 HMT Set1 (Roguev et al, 2003) and H3K9 KDM Lsd1 to euchromatin and to counteract gene silencing (Li et al, 2008).

The *C.elegans* KDM5 homologue rbr-2 was identified as a H3K4me₃ demethylase required for the vulva development (Christensen et al, 2007). Interestingly, rbr-2 can antagonize the role of H3K4 HMT ASH-2 complex in the germ line to modulate worm longevity. Rbr-2 mutation decreased the normal worm life span with a concomitant

increase of H3K4me3 (Greer et al, 2010). Transgenerational epigenetic inheritance of longevity by ASH-2 complex mutation requires *rbr-2* (Greer et al, 2011). These studies suggest unexpected roles of KDM5 family members in epigenetic inheritance, longevity, and possibly stem cell aging.

Consistent with a function for *rbr-2* in regulating life span of *C.elegans*, mutation of the KDM5 *Drosophila* homologue *lid* also produces short-life phenotype in males (Li et al, 2010). *Lid* is encoded by the *little imaginal discs (lid)* gene and genetically categorized as a trithorax-group protein (TrxG) (Gildea et al, 2000). Flies with homozygous *lid* mutations display variable phenotypes: some die in the embryo stage whereas others can survive until pupal development (Gildea et al, 2000). These phenotypes correlate with a global increase of H3K4me3 (Eissenberg et al, 2007; Lee et al, 2007b; Secombe et al, 2007). *Lid* positively regulates the Hox gene-*Ubx* expression (Lee et al, 2007b), is required for dMyc-induced expression of growth regulatory gene *Nop60B* (Secombe et al, 2007), and contributes to the establishment of transcriptionally competent chromatin because *Lid* is required for H3 acetylation and can antagonize heterochromatin-mediated gene silencing (Lloret-Llinares et al, 2008).

Biochemical purification of the *Lid* complex from *Drosophila* embryos revealed that *Lid* stably associates with the HDAC Rpd3 and chromodomain-containing protein MRG15 (Lee et al, 2009b; Moshkin et al, 2009). Incorporation of *Lid* into this complex does not affect *Lid* demethylase activity whereas the Rpd3 histone deacetylase activity is greatly diminished, providing an indirect mechanism explaining why *Lid* is classified as a TrxG protein. MRG15 can specifically recognize H3K36me2/me3, a gene body associated histone modification marking actively transcribed chromatin (Barski et al, 2007; Lee & Shilatifard, 2007; Zhang et al, 2006). Importantly, MRG15 binding can

demarcate a distinct chromatin architecture that is specifically enriched for H3K36me3 and encodes active genes with highly specified biological functions (Filion et al, 2010). It is conceivable that the MRG15-Lid complex may also occupy gene bodies of actively transcribed chromatin via association with H3K36me3 and regulate the expression of distinct categories of active genes. In addition, Lid complex purification also pulls down the Mcm complex components (Moshkin et al, 2009), suggesting that Lid may also regulate DNA replication.

Mammalian KDM5 family homologues KDM5A-D all harbor active H3K4me3 demethylase activity (Christensen et al, 2007; Klose et al, 2007; Lee et al, 2007a; Yamane et al, 2007). *KDM5C* and *KDM5D* localize on the X and Y chromosome, respectively and lack the C-terminal PHD2/3 domain. *KDM5C* mutation is involved in X-linked mental retardation (Tzschach et al, 2006) and the REST complex can recruit KDM5C at the promoters of non-neuronal genes to repress gene expression through demethylating H3K4me3 (Tahiliani et al, 2007). The function of *KDM5D* is poorly understood. In contrast, *KDM5A* or *KDM5B* promotes stem cell phenotype or oncogenic activity (discussed below) while *KDM5C* appears to be a tumor suppressor in cervical cancers (Smith et al, 2010) and clear cell renal carcinomas (Dalglish et al, 2010).

KDM5A/Rbp2 mRNA is ubiquitously expressed and was originally discovered to interact with the tumor suppressor retinoblastoma protein (pRb), which is involved in the cell cycle regulation, cellular senescence and differentiation (Fattaey et al, 1993). KDM5A appears to have nucleolar localization in fibroblasts and can repress the tumor suppressor genes (p21, p27) or activate transcription of homeotic genes (BRD2 and BRD8) depending on the context (Benevolenskaya et al, 2005). *Kdm5a* null mice are viable and display minor haematopoiesis defects (Klose et al, 2007). Genome-wide

ChIP-on-chip analysis of KDM5A in human U937 cells revealed that KDM5A associates with a transcriptional program involved in differentiation (Lopez-Bigas et al, 2008). Mouse *Kdm5a* can target Hox genes in mouse ESCs and suppress their expression via demethylating H3K4me3/me2. During retinoid acid (RA) induced ESC differentiation, *Kdm5a* occupancy decreases at Hox gene promoters and permits their expression associated with an increased H3K4me3 (Christensen et al, 2007). In addition, *Kdm5a* forms a complex with PcG proteins and share many overlapping target genes in ESCs. PcG can recruit *Kdm5a* at lineage-associated gene promoters in ESCs to fine-tune the balance of H3K4me3 and H3K27me3 to repress their expression, and poise for developmental induction (Pasini et al, 2008). Importantly, KDM5A is also part of the MRG15 complex that associates with elongating RNA Pol II and removes intragenic H3K4me3 of actively transcribed genes in human HeLa cells (Hayakawa et al, 2007), suggesting that KDM5A can localize at both promoters and gene bodies depending on differential incorporation into distinct protein complexes. Therefore, KDM5A may function as both a repressor and an activator of transcription in a context dependent manner.

KDM5B/PLU-1 was originally discovered by virtue of its differential expression in response to HER2 signaling perturbation in a mammary epithelial cell line (Lu et al, 1999). Mouse *Kdm5b* is highly enriched in the E5.5 mouse epiblast--the founder tissue of embryo proper but its expression quickly diminishes at E6.5 (Frankenberg et al, 2007). KDM5B protein localizes to the nucleus and its expression remains low in normal adult tissues except for the testis, a reservoir for pluripotent cells (Barrett et al, 2002; Lu et al, 1999). *Kdm5b* knockout mouse embryos are peri-implantation lethal (Catchpole et al, 2011) and *Kdm5b* mutant mouse ESCs cannot be derived from knockout embryos (Table 1.2). This suggests that *Kdm5b* has non-redundant function in regulating early embryogenesis and specifying pluripotent stem cells compared to its close homologue

Kdm5a. In addition, Dey and colleagues failed to generate stable *Kdm5b* shRNA knockdown mouse ESCs (Dey et al, 2008). In contrast, Schmitz et al. reported that knockdown of *Kdm5b* in mouse ESCs did not affect ESC growth or self-renewal (Schmitz et al, 2011). Moreover, they generated *Kdm5b* conditional knockout ESCs and reported that Cre-mediated deletion of *Kdm5b* did not affect ESC self-renewal (Schmitz et al, 2011). Because they did not characterize the transcriptome or the developmental potential of the *Kdm5b* null ESCs, it is not clear whether the *Kdm5b* null ESCs are developmentally normal. Because *Nanog* null ESCs could be generated *in vitro*, it is possible that *Kdm5b* is not strictly required for ESC maintenance but important for the establishment of the pluripotent state similar to *Nanog*. The resolution of these discrepancies merits further investigation. A detailed discussion of these observed discrepancies can be found in the Discussion chapter.

Nanog		+/+	+/-	-/-
Born mice		80	151	0
7.5dpc		14	16	6 (no decidua)
5.5dpc		N.A.	N.A (+/+ plus +/- =26)	7 (no discernable epiblast or ExEcto)
3.5dpc		N.A.	N.A (+/+ plus +/- =12)	3 (normal embryo morphology) ICM of null blastocyst failed to proliferate; Immunosurgical ICM do not persist as undifferentiated mass.

Kdm5b		+/+	+/-	-/-
Born mice	1A4	67	72	0
	1A8	57	49	0
7.5dpc	1A4	0	7	0
	1A8	1	13	0
4.5dpc		N.A.	N.A	Embryo survive, but fail to derive KDM5B ^{-/-} ESCs.

Table 1.2 A comparison between Kdm5b and Nanog homozygous mutant embryo.

Data was adapted from (Catchpole et al, 2011; Mitsui et al, 2003). 1A4 and 1A8 represent two ESCs clones positive for the integration of the Kdm5b targeting vectors.

On the other hand, forced expression of Kdm5b increases the ESC mitotic rate and sustains proliferating progenitors while reducing terminally differentiating cells during differentiation (Dey et al, 2008), suggesting Kdm5b's role in cell fate decision between proliferation and differentiation. We will examine the detailed molecular mechanisms of Kdm5b in regulating ESC proliferation and pluripotency in the following chapters. We also studied the role of zebrafish Kdm5b in early embryogenesis and found that depletion of a single zebrafish Kdm5b paralogue affects gross growth rate and leads to lethality 72 hrs after morpholino microinjection (Appendix Figure 1).

4. KDM5 family members and cancer

Both KDM5A and KDM5B have been implicated in tumorigenesis and tumor progression. KDM5A fusion with nuclear pore complex protein NUP98 induces acute myeloid leukemia (AML) (van Zutven et al, 2006). KDM5A-NUP98 fusion proteins retain the PHD3 of KDM5A, which binds H3K4me3 of Hox genes involved in leukemogenesis and prevents their silencing by PcG (Wang et al, 2009a). Interestingly, increased *KDM5A* expression also confers a drug resistance phenotype while knockdown of *KDM5A* renders cancer cells drug sensitive (Sharma et al, 2010). KDM5B protein was originally found to be upregulated in breast cancers (Lu et al, 1999). Knockdown of *KDM5B* derepresses the expression of a few tumor suppressor genes in breast cancer cells and suppresses tumor growth using a mouse syngeneic cancer model (Yamane et al, 2007). KDM5B can also suppress the angiogenic and metastatic potential of breast cancer cells *in vivo* by repressing chemokine CCL14 (Li et al, 2011). Expression of

KDM5B mRNA is increased in prostate cancers and correlates with the degree of prostate cancer progression. *KDM5B* was shown to interact with androgen receptor (AR) and activates the AR target gene PSA (Xiang et al, 2007). Additionally, *KDM5B* is upregulated in bladder cancer, lung cancer, leukemia and was proposed to target the E2F1/E2F2 genes (Hayami et al, 2010). Interestingly, Roesch and colleagues found that increased *KDM5B* expression marks a small population of slow cycling melanoma cells, which display stem cell properties with self-renewal potential (Roesch et al, 2010). *KDM5B* depletion severely impairs the long-term growth and self-renewal properties of these cells, suggesting that *KDM5B* marks the stemness of melanoma model. These studies suggest that *KDM5A* and *KDM5B* are promising drug targets for cancer therapy.

Interestingly, many *KDM5* family members complex with HDACs (Barrett et al, 2007; Hayakawa et al, 2007; Lee et al, 2009b; Tahiliani et al, 2007) and in the case of *KDM5A*-mediated anti-cancer drug resistance, treatment with HDAC inhibitors can reverse the drug resistance phenotype (Sharma et al, 2010). The *KDM5* fission yeast homologue *Msc1* coprecipitates with histone deacetylase activity. Cells lacking *Msc1* increase the global H3 acetylation level and are sensitive to HDAC inhibitor trichostatin A (TSA) (Ahmed et al, 2004). A recent study reveals that cells treated with HDAC inhibitors display a global increase in H3K4me3 and decreased expression of *KDM5* family members (Huang et al, 2010), suggesting that *KDM5* and HDAC activity are intrinsically connected. Because HDACs inhibitors have been widely used in clinic to treat human cancers (Marks & Breslow, 2007), developing small molecules that specifically target *KDM5* family members may hold important therapeutic values. In addition, *KDM5* family enzymes display more stringent specificity for histone substrates than HDACs. Therefore, *KDM5*-family specific inhibitors may have fewer side effects than HDAC inhibitors.

CHAPTER TWO

KDM5B Regulates Embryonic Stem Cell Self-Renewal by Repressing Cryptic Intragenic Transcription

(Xie et al. 2011, published on EMBO J. 2011 Apr 20;30(8):1473-84.)

**KDM5B Regulates Embryonic Stem Cell Self-Renewal by Repressing Cryptic
Intragenic Transcription**

Liangqi Xie, Carl Pelz, Wensi Wang, Amir Bashar, Olga Varlamova, Sean Shadle, Soren
Impey*

Oregon Stem Cell Center, Papé Family Pediatric Research Institute, Cell and
Developmental Biology, Oregon Health and Science University.

*Correspondence: Soren Impey, Oregon Stem Cell Center, Oregon Health and Science
University, 3181 SW Sam Jackson Park Rd./ L321, Portland OR 97239-3098,
impeys@ohsu.edu, (Phone:503-494-7540, Fax:503-494-5044)

Abstract

Although regulation of histone methylation is believed to contribute to embryonic stem cell (ESC) self-renewal, the mechanisms remain obscure. We show here that the histone 3 trimethyl lysine 4 (H3K4me3) demethylase, KDM5B, is a downstream Nanog target and critical for ESC self-renewal. Although KDM5B is believed to function as a promoter-bound repressor, we find that it paradoxically functions as an activator of a gene network associated with self-renewal. ChIP-Seq reveals that KDM5B is predominantly targeted to intragenic regions and that it is recruited to H3K36me3 via an interaction with the chromodomain protein MRG15. Depletion of KDM5B or MRG15 increases intragenic H3K4me3, increases cryptic intragenic transcription, and inhibits transcriptional elongation of KDM5B target genes. We propose that KDM5B activates self-renewal-associated gene expression by repressing cryptic initiation and maintaining an H3K4me3 gradient important for productive transcriptional elongation.

Subject Categories: Chromatin & Transcription; Genomic & Computational Biology

Keywords: chromatin; epigenetics; histone demethylase; self-renewal; transcriptional elongation

Introduction

ESCs derived from the inner cell mass (ICM) of pre-implantation embryos can differentiate into all somatic lineages and have an unlimited capacity for self-renewal (Niwa, 2007a). ESCs serve as a model for the study of development specification and as an important resource for cell replacement therapy. Insight into the transcriptional and epigenetic mechanisms that regulate ESC self-renewal is essential for their use in medicine.

The transcription factors Oct4 and Nanog play a central role in the initiation and maintenance of ESC pluripotency and self-renewal (Chambers et al, 2003; Do & Scholer, 2009; Mitsui et al, 2003). Nanog is also essential for the acquisition of pluripotency in the embryo and during reprogramming (Silva et al, 2009). ESC pluripotency is believed to depend on an epigenetic state that facilitates unlimited self-renewal and prevents developmental gene silencing. ESCs are characterized by a relaxed chromatin conformation (Niwa, 2007b) and bivalent epigenetic domains where the transcription-associated H3K4me3 mark accumulates repressive H3K27me3 (Bernstein et al, 2007). During reprogramming of somatic cells, extensive epigenetic remodeling occurs, including loss of repressive histone and DNA methylation and re-establishment of bivalent histone modifications (Maherali et al, 2007). Although Nanog has been reported to interact with chromatin complexes (Liang et al, 2008), the downstream mechanisms by which Nanog regulates initiation and maintenance of pluripotency remain unclear. We utilized a ChIP-Seq screen to identify the histone demethylase gene *Kdm5b* as a major downstream target of Nanog.

KDM5B is an H3K4me2/3 demethylase that is upregulated in a wide range of human cancers and enhances cancer self-renewal (Hayami et al, 2010; Lu et al, 1999; Roesch

et al, 2010; Yamane et al, 2007). Forced *Kdm5b* expression also increases histone 3 Ser10 phosphorylation in ESC and reduces neural progenitor differentiation (Dey et al, 2008). Nevertheless, the molecular mechanisms by which *Kdm5b* regulates self-renewal have not been well characterized. Although KDM5B is believed to function as a transcriptional repressor by removing promoter associated H3K4me3, we show here that KDM5B also functions as an activator of self-renewal-associated gene expression. ChIP-Seq analysis shows that KDM5B occupancy is highly correlated with H3K36me3, a chromatin mark associated with transcriptional elongation. KDM5B is recruited to intragenic H3K36me3 at least, in part, via its interaction with the chromodomain MRG15. Moreover, we demonstrate that KDM5B safeguards transcriptional elongation by repressing spurious intragenic transcription. Our study reveals that compartmentalization of a histone demethylase to a distinct chromatin domain results in an unexpected role in gene regulation.

Results

***Kdm5b* is a transcriptional target of Nanog and contributes to ESC self-renewal**

Nanog and Oct4 are core transcriptional regulators of ESC self-renewal. Analysis of ChIP-Seq data showed that Nanog and Oct4 co-occupy the *Kdm5b* genomic locus in mouse ESCs (Figure 2.1A). ChIP-PCR confirmed Nanog and Oct4 occupancy at these regions in both mouse and human ESCs (Figure 2.1B; Supplementary Figure 2.1). Tetracycline-mediated depletion of Oct4 in ZHBTc4 ESCs (Niwa et al, 2000) markedly reduced *Kdm5b* protein and mRNA levels (Figure 2.1C). Likewise, depletion of endogenous Nanog or Oct4 in ESCs attenuated *Kdm5b* expression (Figure 2.1D). Transient knockdown of *Kdm5b* increased expression of lineage-associated genes (Figure 2.1E). These genes are likely indirectly regulated because the majority of them are not associated with KDM5B occupancy as assessed by ChIP-Seq. Prolonged knockdown of *Kdm5b* triggered morphological differentiation, loss of alkaline phosphatase activity, and a reduction in a subset of pluripotency-associated genes (Figure 2.1F; Supplementary Figure 2.1B). Consistent with its proposed role in cancer stem cell self-renewal (Roesch et al, 2010), *Kdm5b* depletion markedly decreased ESC proliferation (Supplementary Figure 2.2A). In particular, S phase was reduced and the percentage of cells in G1 increased (Supplementary Figure 2.2B). These data suggest that *Kdm5b* contributes to ESC self-renewal.

KDM5B is an activator of self-renewal-associated gene expression

Previous studies suggest that KDM5 family demethylases repress genes involved in developmental processes (Christensen et al, 2007; Dey et al, 2008; Lopez-

Bigas et al, 2008; Pasini et al, 2008). We explored the mechanisms by which KDM5B regulates ESC self-renewal by profiling gene expression following siRNA-mediated *Kdm5b* knockdown. Surprisingly, the majority of significantly-regulated genes showed decreased expression. Moreover, Gene ontology (GO) analyses revealed that the most significantly-regulated categories were exclusively associated with genes down-regulated by *Kdm5b* knockdown (Figure 2.2A). Significant GO categories were not associated with developmental processes but were linked to mitosis, chromatin, and nucleotide metabolism. Interestingly, genes downregulated by *Kdm5b* depletion were significantly (KS test $p < 1 \times 10^{-6}$) correlated with genes whose expression decreases during ESC differentiation (Figure 2.2B). Real-time PCR confirmed the downregulation of a subset of these genes following *Kdm5b* knockdown (Supplementary Figure 2.3). These data strongly suggest that KDM5B functions as an activator of many genes associated with self-renewal.

KDM5B occupies transcribed regions of self-renewal associated genes

The ability of KDM5B to function as an activator can result from indirect regulation. To address this question we utilized ChIP-Seq to profile the genome-wide occupancy of KDM5B in an unbiased manner. 8,425,313 ChIP-Seq tags were sequenced of which 3,864,894 mapped to a single genomic locus. A sliding-window algorithm identified 11,142 KDM5B ChIP-Seq peaks significantly enriched at an FDR of 5%. The KDM5B antibody used for ChIP-Seq identified only one protein by western blot (Supplementary Figure 2.4A) and ChIP-PCR confirmed occupancy of a randomly-selected subset of peaks with antibodies to distinct epitopes (Supplementary Figure 2.4B and 2.4C). Moreover, siRNA-mediated depletion of *Kdm5b* attenuated KDM5B ChIP signal (Figure

2.2C). GO analysis of genes adjacent to KDM5B ChIP-Seq peaks generated categories that were highly similar to those associated with genes downregulated following *Kdm5b* knockdown (Figure 2.2A and 2.2D). Surprisingly, unbiased ChIP-Seq revealed that 84% of KDM5B ChIP-Seq peaks were present in intragenic regions (Wilcoxon Rank-Sum $p < 1 \times 10^{-6}$; Figure 2.2E-G) while only 36% of random loci and 40% of Oct4 ChIP-Seq peaks were located inside genes.

The accumulation of KDM5B in transcribed regions suggested an association with transcriptional activity. We utilized ChIP-Seq to profile occupancy of a phosphorylated form of Pol II associated with elongation (Ser2P) and the promoter-associated epigenetic mark, H3K4me3. We found that ~59% of KDM5B peaks were within 2 kb of a Ser2P peak (Wilcoxon Rank-Sum $p < 1 \times 10^{-6}$; Figure 2E-G) while only 6% of random loci and 13% of Oct4 ChIP-Seq peaks were adjacent to a Ser2P peak. KDM5B was not significantly associated with H3K4me3 (Wilcoxon Rank-Sum $p = 0.23$). These data strongly suggest that KDM5B functions as a transcriptional activator. Consistent with this idea KDM5B ChIP-Seq tags accumulated in intragenic regions of genes positively-regulated by KDM5B but not in intragenic regions of genes repressed by KDM5B (Figure 2.2H and 2.2I). Comparison of ChIP-Seq tag accumulation relative to ranked microarray data reveals that KDM5B occupancy is significantly correlated with genes decreased by *Kdm5b* knockdown (Figure 2.2J; Supplemental Figure 2.5). Consistent with this finding, KDM5B occupies intragenic regions of highly expressed RefSeq genes but was not significantly detected in genes expressed at lower levels (Figure 2.2K-M). The lack of ChIP-Seq peak correlation with genes increased following *Kdm5b* knockdown shows that KDM5B functions predominantly as an activator of gene expression. Both microarray and ChIP-Seq studies identify an overlapping set of genes that are transcriptionally activated by KDM5B. Taken together, these data indicated that KDM5B

regulates self-renewal by functioning as a direct activator of a self-renewal-associated gene network. Analyses of microarray and ChIP-Seq data identified a large set of KDM5B targets associated with cell cycle progression and DNA biosynthesis, including *Ccnb1*, *Ccna2*, *Cdc25a*, *Pola1*, *Mcm3/4/5/6/7*, *Cdc45*, and *Orcl1/2/3/5/6* (Supplemental Figure 2.6A). We confirmed that these genes were directly occupied by KDM5B (Supplemental Figure 2.6B). We also show here that a subset of KDM5B target genes directly regulates proliferation and DNA synthesis (Supplemental Figure 2.6B). These data indicate that KDM5B regulates a novel gene network that directly regulates ESC self-renewal.

Demethylation of intragenic H3K4me3 domains by KDM5B

Although the majority of H3K4me3 is localized near promoters (Mikkelsen et al, 2007), our data indicates that KDM5B is predominantly localized to intragenic regions. This led us to examine whether KDM5B functions as a demethylase at intragenic loci. Knockdown of *Kdm5b* modestly increased global H3K4me3, but not H3K9me3, H3K27me3 or H3K36me3 in ESCs (Figure 2.3A). To test whether KDM5B specifically regulates intragenic H3K4me3 at its targets, we utilized ChIP-Seq to examine global changes in H3K4me3 following knockdown of *Kdm5b*. Interestingly, we observed highly localized increases in H3K4me3 near KDM5B peaks following knockdown of *Kdm5b* (Figure 2.3B and 2.3C). Importantly, the increase in H3K4me3 at KDM5B ChIP-Seq peaks was highly significant (Wilcoxon rank-sum $p < 1 \times 10^{-6}$) and occurred on a genome-wide basis (Figure 2.3D). We utilized ChIP-PCR to confirm this observation by showing that *Kdm5b* knockdown increased H3K4me3 at ChIP-Seq peaks but not at upstream or

downstream regions (Figure 2.3E). These data show that KDM5B functions to erase intragenic H3K4me3 specifically at its target sites.

Recruitment of KDM5B to chromatin depends on H3K36me3

The enrichment of KDM5B in transcriptionally active regions is reminiscent of the distribution of H3K36me3, an epigenetic mark deposited by elongating Pol II (Li et al, 2007a). We compared the localization of KDM5B to H3K4me3 ChIP-Seq data generated in this study and published H3K36me3 and H3K27me3 ChIP-Seq data (Mikkelsen et al, 2007) (Figure 2.4A and 2.4B). 83% of KDM5B regions were within 5 kb of H3K36me3 peak (Wilcoxon rank-sum $p < 1 \times 10^{-6}$) while only ~18% of KDM5B peaks were near H3K4me3 peaks (Wilcoxon rank-sum $p = 0.23$). Because the majority of KDM5B ChIP-Seq peaks (84%) are localized to RefSeq intragenic regions, we examined the density of H3K36me3 relative to RefSeq genes ordered by KDM5B ChIP-Seq peak area. H3K36me3 density was highly correlated with genes containing the most KDM5B occupancy (Figure 2.4C). Importantly, the spatial profile of KDM5B occupancy in highly expressed genes mirrored that of H3K36me3 (Figure 2.4D and 2.4E). ChIP-PCR confirmed that H3K36me3 was specifically enriched at KDM5B ChIP-Seq peaks but not at down- or upstream regions (Figure 2.4F). The colocalization of KDM5B and H3K36me3 raised the possibility that KDM5B is directly recruited by this epigenetic mark. We tested this by examining KDM5B occupancy following knockdown of the Setd2 H3K36me3 methyltransferases (Figure 2.4G). Setd2 depletion markedly decreased KDM5B occupancy (Figure 2.4H) and induced a localized increase in H3K4me3 (Supplemental Figure 2.7A). These data show that H3K36me3 deposition facilitates recruitment of KDM5B. Because H3K36me3 is deposited by elongating Pol II, KDM5B

recruitment should also be dependent on transcriptional elongation. Inhibition of Pol II elongation by treatment of pTEFb inhibitor DRB markedly attenuated occupancy of KDM5B and elongating Pol II while sparing promoter-bound Pol II (Supplemental Figure 2.7B and 2.7C). These data strongly suggest that KDM5B is recruited by elongation-associated H3K36 methylation.

An Rpd3S-like complex recruits KDM5B to H3K36me3

The *Kdm5b* ortholog, *little imaginal discs* (*Lid*), interacts with the chromodomain H3K36me3-binding protein MRG15 in *Drosophila* (Lee et al, 2009b). Interestingly, mammalian KDM5B also interacts with MRG15 in ESCs (Figure 2.5A). Similar results were seen with epitope-tagged KDM5B and MRG15 in a heterologous expression system (Figure 2.5B). Moreover, real-time ChIP analysis showed that MRG15 was selectively enriched at KDM5B peaks but not at distal regions (Figure 2.5C). Knockdown of MRG15 modestly decreased KDM5B occupancy, increased intragenic H3K4me3, and reduced expression of KDM5B target genes (Figure 2.5D-G). These data indicate that MRG15 contributes to recruitment of KDM5B to H3K36me3. The yeast MRG15 ortholog Eaf3 functions to recruit the Rpd3S histone deacetylase and Sin3 to transcriptionally active chromatin. Thus, we asked whether KDM5B interacts with a mammalian Rpd3S-like complex. Pull down experiments show that *Kdm5b* interacts with mammalian Rpd3S-like complex components HDAC1 and Sin3A (Figure 2.5H). KDM5B was not detected in HDAC1-containing Mi-2/NURD immunoprecipitates. Importantly, *Kdm5b* knockdown did not affect H4 acetylation or H3K36me3 (Supplemental Figure 2.8A). We next utilized ChIP-Seq to test whether MRG15 colocalized with KDM5B on a genome wide scale. MRG15 ChIP-Seq signal was highly correlated with KDM5B occupancy

(Figure 2.5I and 2.5J) and 49% of KDM5B ChIP-Seq peaks were within 5kb of an MRG15 peak (Wilcoxon rank-sum $p < 1 \times 10^{-6}$). Like KDM5B, MRG15 was predominantly associated with intragenic transcription (71% of MRG15 peaks are inside RefSeq genes; Wilcoxon rank-sum $p < 1 \times 10^{-6}$) and the spatial profile of MRG15 occupancy overlapped with that of KDM5B and H3K36me3 (Supplemental Figure 2.8B-C). Moreover, MRG15 occupancy was highly correlated with KDM5B but not Nanog intragenic occupancy (Figure 2.5K). Although the Nanog negative control showed no correlation with KDM5B occupancy, a general enrichment for Nanog (over regions without KDM5B) is expected due to the tendency of KDM5B to demarcate transcriptionally active chromatin. These data strongly suggest that KDM5B can be recruited to H3K36me3 via its interaction with MRG15. Because not all KDM5B peaks are associated with MRG15 occupancy additional recruitment mechanisms likely also exist.

KDM5B represses cryptic transcription by removing intragenic H3K4me3

Rpd3, the catalytic core of the Rpd3S complex, is believed to repress cryptic transcription by erasing histone acetylation deposited by elongating polymerase (Carrozza et al, 2005; Joshi & Struhl, 2005; Keogh et al, 2005). Our observation that KDM5B interacts with mammalian orthologs of Rpd3S complex suggests that KDM5B functions in a manner analogous to Rpd3. In support of this hypothesis we found that mammalian H3K4me3 methyltransferase subunits form a complex with elongating Pol II (Figure 2.6A). Consequently, we examined whether intragenic H3K4me3 deposition was dependent on transcriptional activity. Treatment of ESCs with the inhibitor of Pol II elongation, DRB, significantly decreased intragenic H3K4me3 at both 1hr (Figure 2.6B) and 6hr time points. Unphosphorylated Pol II is tethered to sites of initiation via an

interaction with TFIID/TBP (Nikolov et al, 1995; Usheva et al, 1992). Interestingly, *Kdm5b* knockdown markedly stimulated recruitment of unphosphorylated Pol II to intragenic H3K4me3 peaks strongly suggesting that these sites represent sites of cryptic initiation (Figure 2.6F). Because H3K4me3 is highly correlated with transcriptional start sites and recruits the Pol II pre-initiation complex via interactions with TFIID (Vermeulen et al, 2007), we hypothesized that KDM5B repressed cryptic transcription by preventing intragenic initiation.

Knockdown of *Kdm5b* markedly increased expression of cryptic unspliced transcripts (Figure 2.6C). Moreover, ChIP and global ChIP-Seq analyses show that the increase in cryptic transcription was associated with a localized increase of H3K4me3 (Figure 2.3B-E). These noncoding transcripts did not originate from the canonical promoter because we detected no change in Ser5P Pol II recruitment at promoter regions after *Kdm5b* knockdown (Figure 2.6E). The position of intragenic H3K4me3 at the 3' end of the gene led us to examine whether some of this cryptic transcription was antisense to the coding strand. We found that both *Tet1* and *Unc45b* contained cryptic transcripts antisense to the coding strand (Supplementary Figure 2.9). These cryptic transcripts did not originate from other transcripts because KDM5B-regulated transcription was not detected outside the transcribed locus. Moreover, full-length (intron-spanning) transcription of KDM5B targets were always down-regulated following *Kdm5b* depletion (Figure 2.6D). The decrease in productive (full length and spliced) transcription was associated with a selective decrease in intragenic Ser2P at KDM5B target genes, indicating a defect in later phases of transcriptional elongation (Figure 2.6E). Recruitment of Ser5P to promoter proximal regions was unaffected but intragenic Ser5P also showed decreased recruitment following *Kdm5b* knockdown (Figure 2.6E). These data strongly suggest that KDM5B functions to prevent cryptic transcription by

removing intragenic H3K4me3 (Figure 2.6F) and that KDM5B safeguards expression of a gene network associated with self-renewal by maintaining an H3K4me3 gradient that favors productive elongation (Figure 2.6G).

Discussion

KDM5 family demethylases are believed to repress transcription by removing promoter-associated H3K4me3 (Christensen et al, 2007; Klose et al, 2007; Yamane et al, 2007). Surprisingly, we show that KDM5B paradoxically functions as a transcriptional activator of genes linked to ESC self-renewal. ChIP-Seq analyses reveal that KDM5B predominantly occupies actively transcribed chromatin and is recruited to domains enriched for elongating Pol II. KDM5A has been shown to function as an activator or repressor depending on gene context (Benevolenskaya et al, 2005; Chan & Hong, 2001) and apparently occupies promoters and intragenic regions (Christensen et al, 2007; Hayakawa et al, 2007). Interestingly, the *Drosophila* KDM5B ortholog Lid is associated with transcriptionally active chromatin, positively regulates gene expression, and antagonizes heterochromatin formation (Lee et al, 2007b; Lloret-Llinares et al, 2008; Secombe et al, 2007). Thus, KDM5B's role in gene activation may be evolutionarily conserved.

Nanog is required for the initiation and maintenance of pluripotency (Chambers et al, 2003; Mitsui et al, 2003; Silva et al, 2009). We utilized ChIP-Seq data to identify *Kdm5b* as a Nanog target gene. *Kdm5b* depletion attenuated ESC self-renewal and decreased expression of genes associated with mitosis and nuclear metabolism. In particular, KDM5B regulated core DNA biosynthesis machinery, including Pola1, Cdc45, Mcm3-7, Orc1/2/3/5/6, Cdc25a, and Ccnb1. Interestingly, Mcm4 and Mcm6 contribute to rapid early embryonic proliferation (Coue et al, 1996) and both *Mcm4* and *Cdc25a* null embryos fail to expand ICM (Lee et al, 2009a; Shima et al, 2007). Interestingly, many KDM5B-regulated genes are not targets of core pluripotency-associated transcription

factors. Thus, KDM5B regulates ESC self-renewal by activating a novel gene network that regulates ESC proliferation and DNA synthesis.

We show here that KDM5B recruitment depends on H3K36me₃, a histone modification associated with transcriptional elongation (Li et al, 2007a). The KDM5B ortholog Lid interacts with MRG15, a chromodomain protein that recognizes H3K36me_{2/3} (Lee & Shilatifard, 2007; Zhang et al, 2006). We show that KDM5B interacts with MRG15 and that knockdown of MRG15 reduced KDM5B recruitment. Genome-wide MRG15 ChIP-Seq analysis shows a high degree of colocalization with KDM5B. MRG15 is an ortholog of Eaf3, a yeast Rpd3S histone deacetylase complex component (Carrozza et al, 2005; Keogh et al, 2005) and has also been linked to transcriptionally active genes in *Drosophila* (Filion et al, 2010). Interestingly, KDM5B also interacts with Rpd3S complex orthologs, HDAC1 and Sin3A. Similarly, *Drosophila* Lid and mammalian KDM5A also interact with an Eaf3/MRG15, Sin3, and Rpd3/HDAC1 Rpd3S-like complex (Hayakawa et al, 2007; Moshkin et al, 2009). Our data indicates that KDM5B can be tethered to chromatin by H3K36me_{2/3} via association with an Rpd3S-like complex. Because ~17% of KDM5B ChIP-Seq peaks are not adjacent to H3K36me₃, other mechanisms for recruitment may also exist. Interestingly, Setd2 knockdown and inhibition of Pol II elongation induced a more robust decrease in KDM5B occupancy than MRG15 knockdown, suggesting recruitment to H3K36me₃ may involve additional mechanisms. Recent work has identified a family of PWWP domain proteins that also interact with H3K36me₃ in mammalian cells (Vermeulen et al, 2010; Vezzoli et al, 2010).

Recruitment of Rpd3S by H3K36me₃ is believed to prevent cryptic transcription by removal of transcription-associated histone acetylation (Carrozza et al, 2005; Keogh

et al, 2005). Mutation of Set2 or Rpd3S subunits increases cryptic intragenic transcription in *S. cerevisiae* and antisense transcription in *S. pombe* (Carrozza et al, 2005; Keogh et al, 2005; Nicolas et al, 2007). Interestingly, elongating forms of Pol II also interact with the Set1 H3K4 methyltransferase in yeast (Krogan et al, 2003) and cryptic intragenic deposition of H3K4me3 represses gene expression in yeast (Pinskaya et al, 2009). We show that mammalian Pol II interacts with core subunits of the H3K4 methyltransferase and that inhibition of polymerase elongation reduces intragenic H3K4me3 deposition. Because H3K4me3 serves as a nucleation site for Pol II (Vermeulen et al, 2007), we propose that Kdm5b recruitment functions to prevent cryptic transcription. Consistent with this idea, KDM5B knockdown triggered a marked increase in H3K4me3 specifically at intragenic target sites and ChIP-Seq analyses confirmed that *Kdm5b* knockdown is correlated with localized increase in H3K4me3. Importantly, knockdown of *Kdm5b* increased spurious transcription in Kdm5b target genes. Cryptic transcription induced by *Kdm5b* knockdown is associated with a marked decrease in the transcription of functional full length mRNAs. Decreased production of functional transcripts was associated with repression of elongation because Pol II recruitment was specifically reduced in intragenic regions but not at sites of initiation. We also show that some cryptic transcripts specifically originate from the antisense strand. Cryptic transcription can play a role in repressing productive transcription via inhibition of Pol II processivity, altering nucleosomal structure, or post-transcriptional mechanisms, such as, RNAi. The recent observation that MRG15 regulates PTB-dependent alternative splicing also suggests a possible role for KDM5B's regulation of intragenic H3K4me3 in splicing {Luco, 2010 #334}.

Accumulation of H3K4me3 is believed to generate an epigenetic landscape conducive to high levels of transcriptional activity (Kouzarides, 2007; Li et al, 2007a) via

its ability to recruit initiating Pol II (Vermeulen et al, 2007). Our data suggests that active demethylation of intragenic H3K4me3 by KDM5B is another mechanism by which this H3K4me3 gradient is established. Interestingly, the human H3K4me2/me1 demethylase, LSD2, also accumulates in intragenic regions presumably via its interaction with elongating Pol II (Fang et al, 2010). Nevertheless, several lines of evidence suggest that KDM5B and LSD2 have distinct functions. Unlike H3K4me3, LSD2's substrate (H3K4me2/me1) is not associated with recruitment of initiating polymerase. Moreover, KDM5B but not LSD2 interacts with the Rpd3S component HDAC1 while LSD2 but not KDM5B directly interacts with elongating Pol II (Fang et al, 2010). These experiments suggest that LSD2 and KDM5B exist in different complexes and are recruited to active genes via different mechanisms. A recent study also found that the H3K36 demethylase, KDM2A, is recruited to non-methylated CpG islands and maintains low levels of H3K36me at these regions (Blackledge et al, 2010). Because CpG islands are associated with promoter regions, a phenomenon similar to the action of KDM5B may serve to repress accumulation of H3K36me at promoters.

Recent studies suggest that initiation of elongation is a rate-limiting step for ESC transcriptional activity. 30% of human ESC genes undergo transcriptional initiation without evidence of productive elongation (Guenther et al, 2007) and Pol II is paused at the promoters of many ESC genes (Rahl et al, 2010). Our data showing that KDM5B safeguards productive elongation complements data linking c-Myc to initiation of transcriptional elongation (Rahl et al, 2010) and suggests that regulation of elongation may contribute to ESC self-renewal. Regulators of transcriptional elongation and KDM5-family demethylases have also been implicated in cancer. Nup98 fusions to the NSD1 H3K36 methyltransferase and KDM5A induce leukemia (Wang et al, 2007; Wang et al, 2009a). Moreover, *Kdm5b* is upregulated in human cancers (Hayami et al, 2010; Xiang

et al, 2007; Yamane et al, 2007) and contributes to self-renewal of a subpopulation of melanoma cells required for tumor growth (Roesch et al, 2010). It would be interesting to determine whether epigenetic regulation of transcriptional elongation by KDM5B contributes to cancer cell renewal.

Acknowledgements

We thank Ian Chambers and Austin Smith for providing ZHBtc4 ES cells. We also thank Kaoru Tominaga and Olivia M. Pereira-Smith for providing MRG15 anti-sera and HA-tagged MRG15 cDNA. We thank the Oregon Stem Cell Center Flow Cytometry Core for assistance with flow cytometry data.

Figures and Figure legends

Figure 2.1 *Kdm5b* is a Nanog and Oct4 target and critical for ESC self-renewal. (A) UCSC genome browser track depicting Oct4 and Nanog ChIP-Seq occupancy in the vicinity of the murine *Kdm5b* genomic locus. (B) Confirmation of Oct4 and Nanog occupancy at *Kdm5b* (P1 and P2) by ChIP-PCR. The Oct4 enhancer region is a positive control. (C and D) ZHBTc4 ESCs were treated with tetracycline (Tet) for 48hrs to deplete Oct4 (C). J1 ESCs were transfected with siNanog or vector control (D). *Kdm5b* cDNA and protein levels were measured by RT-PCR (left) and immunoblot (right). (E) RT-PCR analysis of lineage-associated marker gene expression following siKdm5b in ESCs. (F) ESCs were transfected with control (siCon) or *Kdm5b* siRNA (siKdm5b). 72 hrs post-transfection cells were stained for alkaline phosphatase activity (AP).

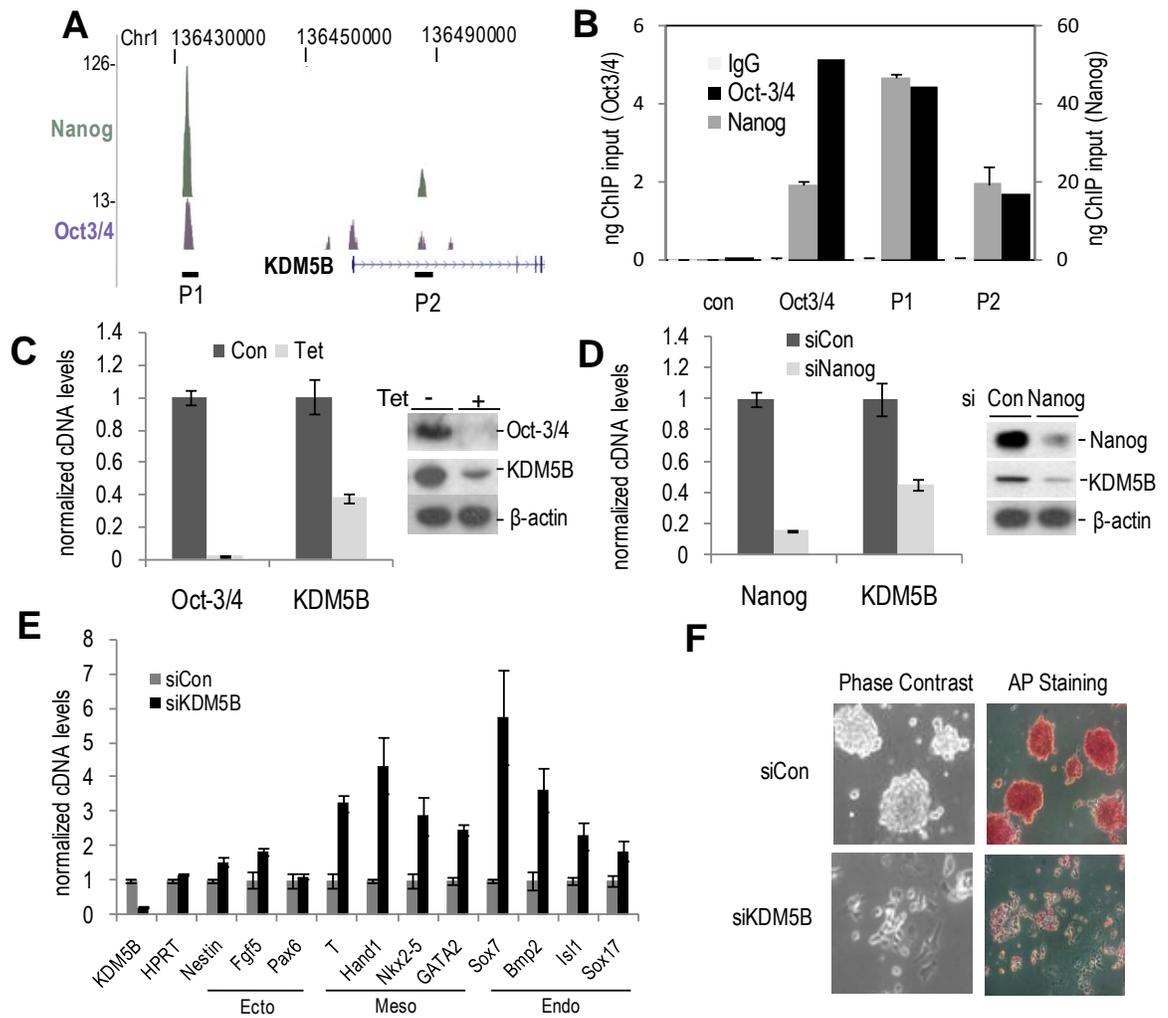


Figure- 1

Figure 2.2 An unexpected role for KDM5B in transcriptional activation. (A) Gene ontology (GO) analysis of microarray data showing categories of genes significantly downregulated following siKdm5b. No upregulated gene categories were detected at $Q < 1 \times 10^{-6}$. (B) Heatmap comparison of microarray data from ESC retinoic acid (RA) differentiation and *Kdm5b* knockdown experiments. (C) ChIP analysis of KDM5B ChIP-Seq peaks and negative control regions following *Kdm5b* knockdown (siKdm5b). Data are expressed as the ratio of siKdm5b ChIP signal over siCon. Each dot represents a distinct ChIP-Seq locus (pairwise comparisons to Peak data $p < 0.01$). (D) GO analysis of RefSeq genes within 5 kb of a *Kdm5b* ChIP-Seq peak 5' end. (E-G) UCSC genome browser tracks depicting H3K4me3, Kdm5b, and Ser 2 phosphorylated Pol II (Ser2P) ChIP-Seq peaks at representative gene loci. The RNA-Seq track depicts 3'-directed RNA-Seq data from ESCs. The numbers on the left axes indicate peak amplitude. (H and I) ChIP-Seq tag density relative to significantly down- (H) and up-regulated (I) genes following *Kdm5b* knockdown. All Ref-Seq gene lengths were scaled to 1. The left Y-axis corresponds to H3K4me3 ChIP-Seq data while the right Y-axis represents other data. (J) Histogram of KDM5B ChIP-Seq peak tag density relative to ESC RefSeq microarray data rank-ordered by fold-change following *Kdm5b* knockdown. The black profile above the heat map depicts average number of significant ChIP-Seq KDM5B peak tags found per gene. The gray profile represents a randomized control and the blue line delineates an FDR of $p < 0.01$. (K-M) ChIP-Seq tag density relative to the top (K), middle (L), and bottom third (M) of RefSeq genes expressed in ESCs. All Ref-Seq gene lengths were scaled to 1. The left Y-axis corresponds to H3K4me3 ChIP-Seq data while the right Y-axis represents other data.

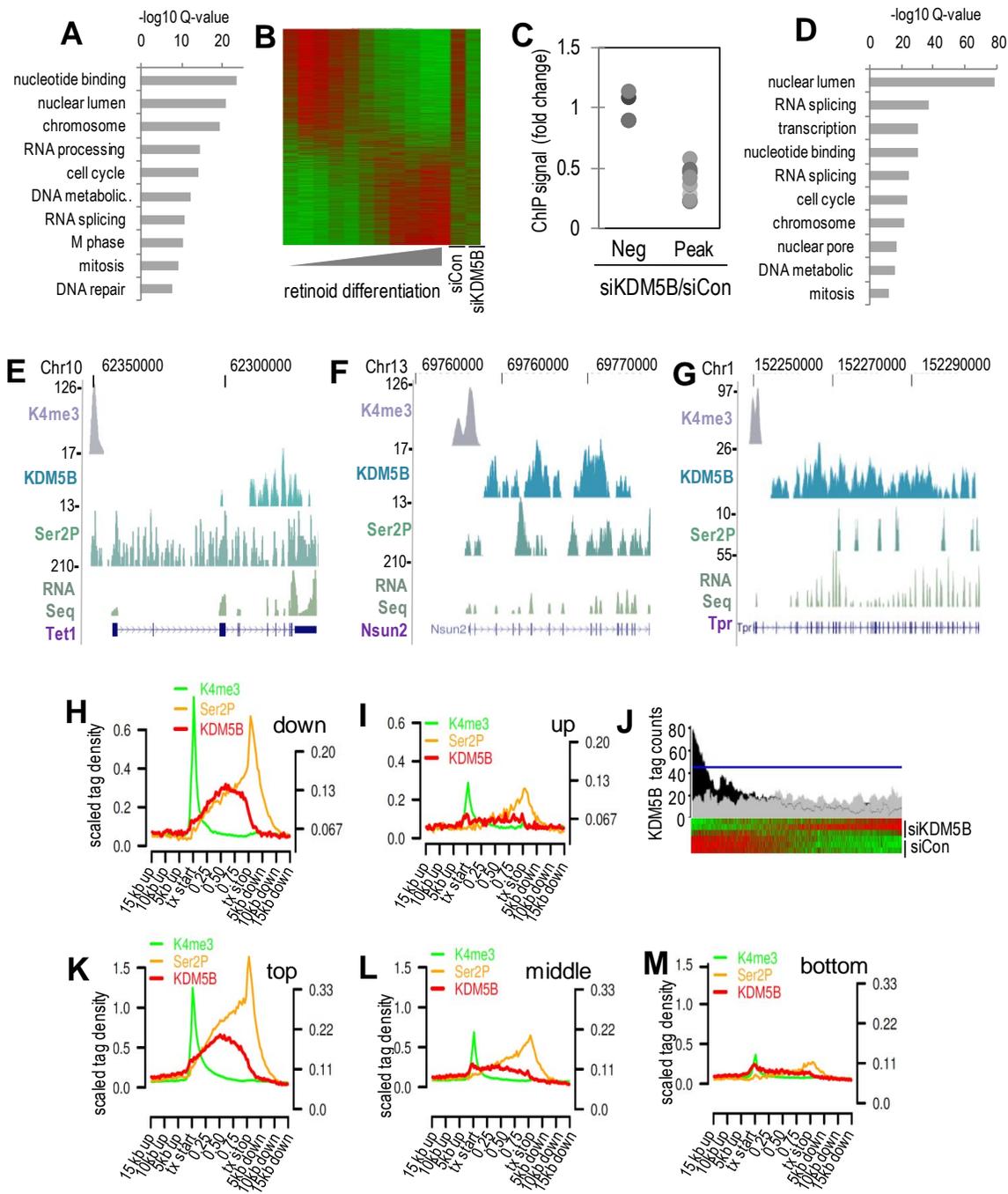


Figure-2

Figure 2.3 KDM5B removes local domains of intragenic H3K4me3. (A) Immunoblot analysis of bulk histone modifications from ESCs transfected control/*Kdm5b* siRNAs . (B-C) Genome browser tracks depicting KDM5B ChIP-Seq peaks and H3K4me3 ChIP-Seq peaks from control (siCon) or *Kdm5b* knockdown cells (siKdm5b). The numbers on the left axes indicate peak amplitude. (D) Histogram depicts the difference of H3K4me3 ChIP-Seq tag counts following knockdown of *Kdm5b* (siKdm5b) versus control (siCon). ChIP-Seq tag counts are plotted relative to the center of mass of KDM5B ChIP-Seq peaks (black line) or randomized controls (gray line). The ChIP-Seq difference was normalized by total tag counts and tags within 1.5 kb of RefSeq gene 5' ends were not included. The difference of H3K4me3 ChIP-Seq tag counts following knockdown of *Kdm5b* was highly significant (Wilcoxon rank-sum $p < 1 \times 10^{-6}$). (E) Scatter plot depicts H3K4me3 ChIP data following *Kdm5b* knockdown (siKdm5b) at KDM5B ChIP-Seq peaks (Peak), upstream regions (Up), associated promoters (Prom), and downstream regions (Down). Data are expressed as the ratio of H3K4me3 ChIP signal from siKdm5b over siCon. Each dot represents a distinct ChIP-Seq peak locus and all comparison to Peak data $p < 0.01$.

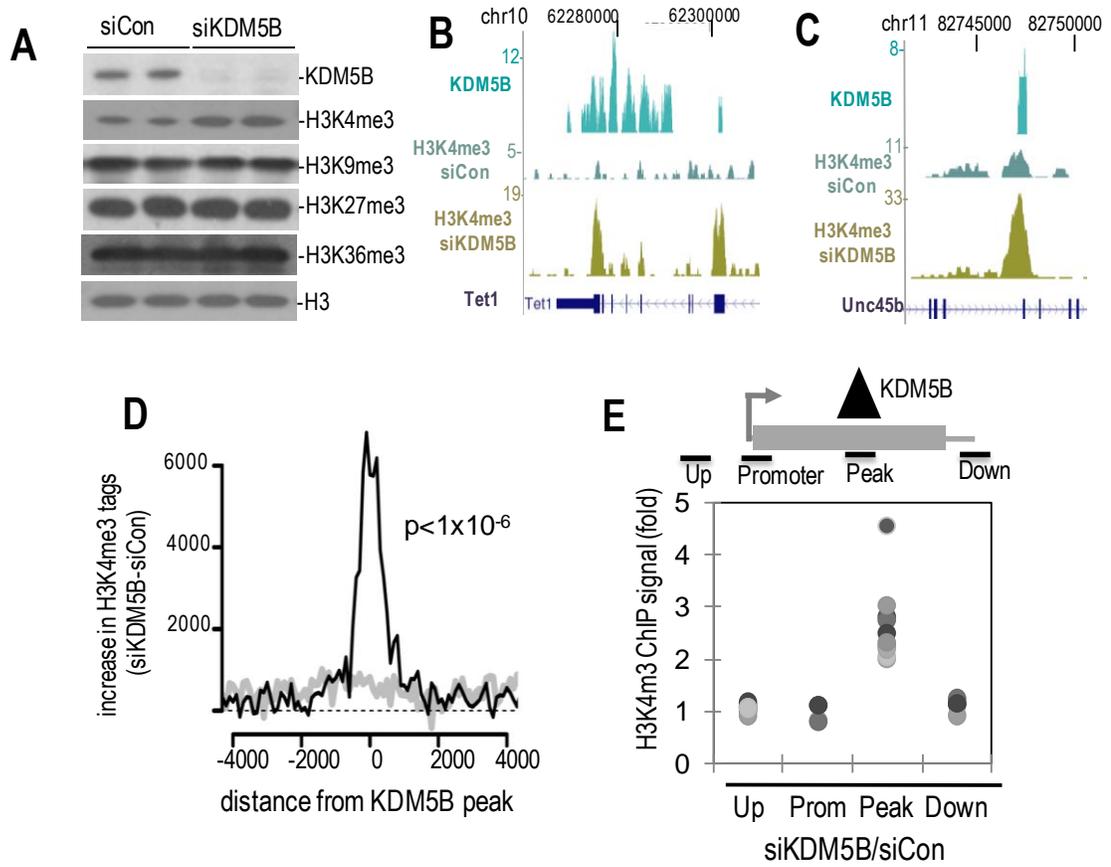


Figure- 3

Figure 2.4 H3K36me3 recruits KDM5B to transcriptionally active intragenic regions. (A and B) UCSC genome browser tracks depicting H3K4me3, KDM5B, and H3K36me3 ChIP-Seq peaks at representative gene loci. The numbers on the left axes indicate peak amplitude. (C) Histogram depicting H3K36me3, and H3K27me3 (K27) ChIP-Seq peak density in RefSeq genes rank-ordered by KDM5B peak area. The gray bar indicates genes without KDM5B peaks (5x scaling). (D and E) ChIP-Seq tag density relative to 5000 highly expressed RefSeq genes (D) and 5000 genes expressed at low levels (E). All Ref-Seq gene lengths were scaled to 1. The left Y-axis corresponds to H3K4me3 ChIP-Seq data while the right Y-axis represents other data. (F) Scatter plot depicts ChIP-PCR for H3K36me3 at KDM5B ChIP-Seq peaks (Peak), and control regions (Up and Down). Each dot represents a distinct ChIP-Seq locus and pairwise comparisons to Peak data $p < 0.01$. (G) Immunoblot of ESC extracts with the indicated antibodies following knocking down H3K36 methyltransferase Setd2. (H) KDM5B ChIP-PCR signal at negative control regions (Up and Down) and KDM5B peaks following knockdown of Setd2 (siSetd2). Data are expressed as the ratio of KDM5B ChIP signal of siSetd2 over siCon. Each dot represents a distinct locus and pairwise comparisons to Peak data $p < 0.01$.

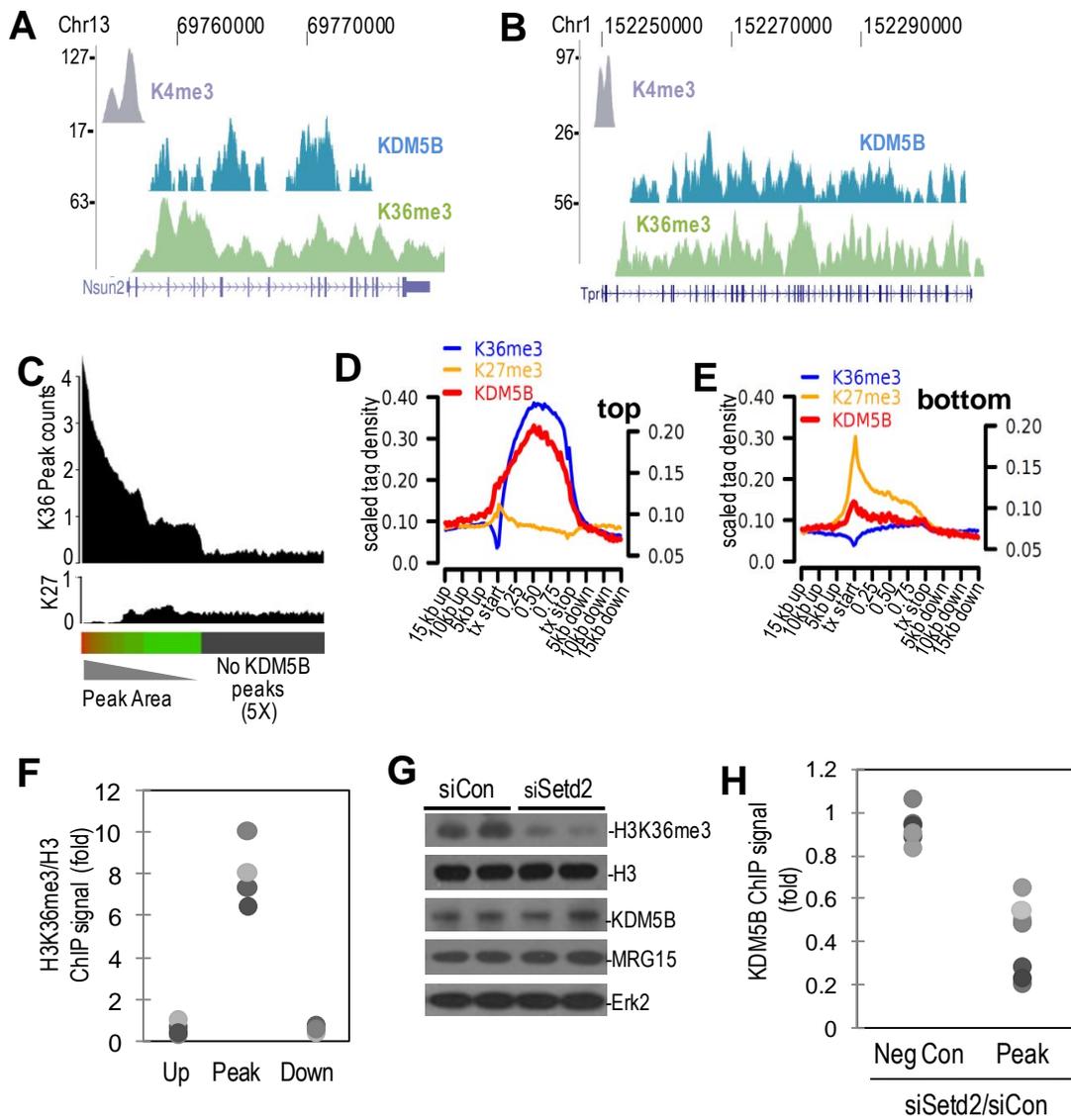


Figure- 4

Figure 2.5 MRG15 mediates recruitment of KDM5B. (A and B) Endogenous (A) or epitope-tagged (B) KDM5B and MRG15 immunoprecipitates from ESCs were immunoblotted with the indicated antibodies.(C) CHIP-PCR analysis of MRG15 occupancy at KDM5B peaks (Peak) and control regions(Up and Down).(D) Immunoblot analysis of KDM5B and MRG15 from ESC whole cell extracts following MRG15 knockdown (siMRG15). (E and F) KDM5B or H3K4me3 CHIP-PCR analysis at negative control regions (Up and Down) and KDM5B peaks following MRG15 knockdown (siMRG15). Data are expressed as the ratio of CHIP signal from siMRG15 over siCon. Each dot represents a distinct CHIP-Seq locus and pair-wise comparisons to Peak data $p < 0.01$. (G) RT-PCR analysis of randomly picked KDM5B target genes (Ccnb1, Mcm4, Fubp1, Tmem48) and control (HPRT, Oct4) following siRNA-mediated MRG15 knockdown. (H) KDM5B, HDAC1 and SIN3A immunoprecipitates from ESCs were immunoblotted with the indicated antibodies. (I and J) UCSC genome browser tracks depict H3K36me3, KDM5B, and MRG15 CHIP-Seq peaks at representative gene loci. The numbers on the left axes indicate peak amplitude. (K) Histogram depicting MRG15 and Nanog (control) CHIP-Seq peak density in RefSeq genes rank-ordered by KDM5B peak area. The gray bar indicates genes without KDM5B peaks (5x scaling).

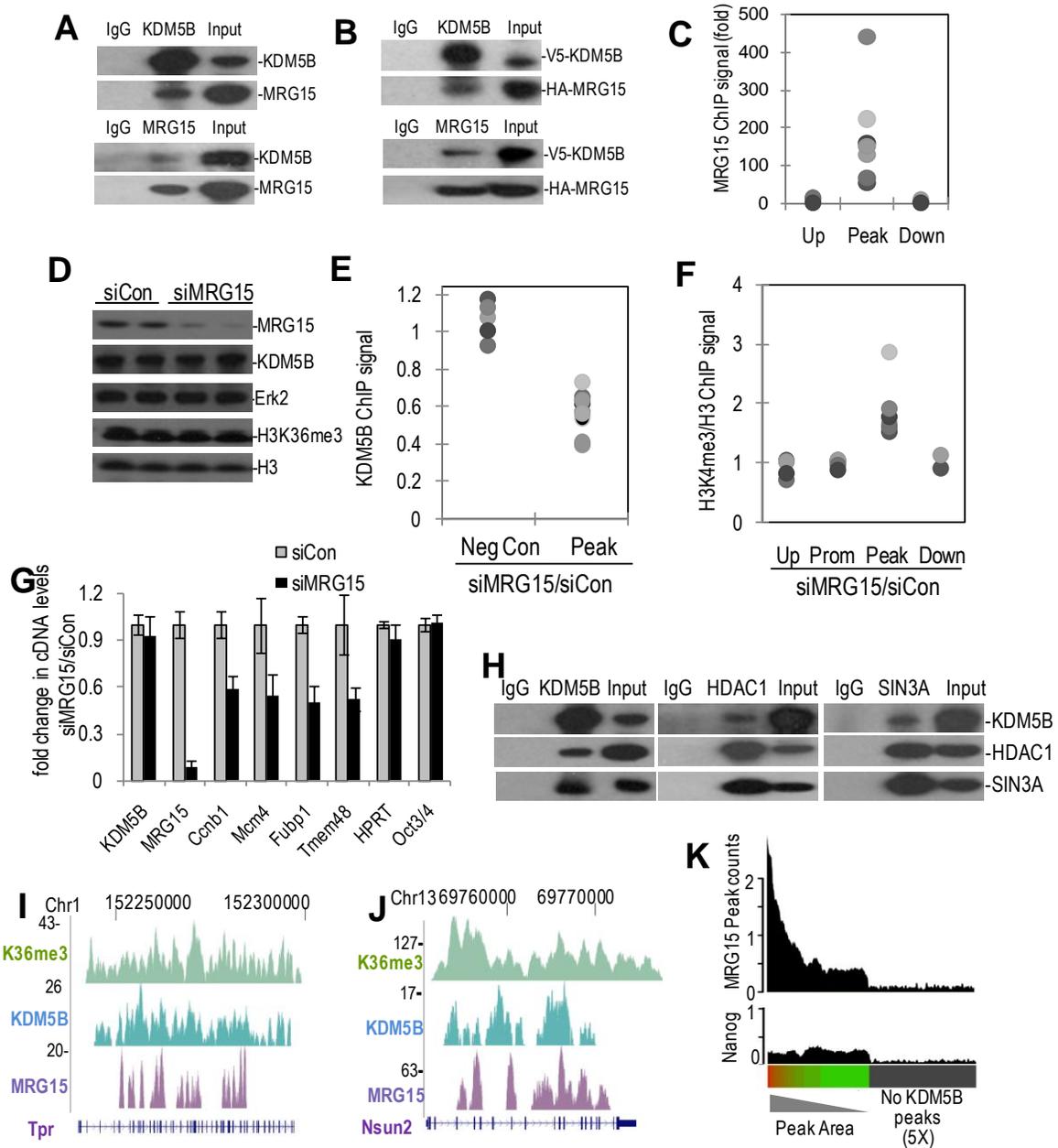


Figure- 5

Figure 2.6 KDM5B safeguards transcriptional elongation by repressing cryptic transcription. (A) Immunoblot analysis of immunoprecipitated WDR5, Ash2L, Ser5P and Ser2P with indicated antibodies. (B) ChIP-PCR analysis of Ser2P, Ser5P, and H3K4me3 at KDM5B ChIP-Seq peaks and associated promoters following 1hr DRB treatment. Data is expressed as fold change versus untreated. Each dot represents a distinct ChIP-Seq locus. (C and D) RT-PCR measurement of cryptic transcription (left panel) and full-length mRNAs (right panel) following knock down of *Kdm5b* (siKdm5b) at negative control genes (*Hprt*, *Oct4*, *Nanog*) and KDM5B target genes (*Ccnb1*, *Cdc25a*, *ATM*, *Mcm4*, *Fubp1*, *Tmem48*, *TET1*, *Unc45b*). (E) ChIP-PCR analysis of Ser5P and Ser2P occupancy at KDM5B ChIP-Seq peaks (Target) and non-target regions. Data are expressed as the ratio of H3K4me3 ChIP signal from siKdm5b over siCon. Each dot represents a distinct ChIP-Seq locus. (F) ChIP analysis of un-phosphorylated Pol II occupancy at control promoter regions and intragenic H3K4me3 peaks following *Kdm5b* knockdown in J1 ESCs. (G) Deposition of intragenic H3K4me3 by Pol II results in initiation of cryptic transcription. The KDM5B demethylase is recruited to H3K36me3 via interaction with an Rpd3S-like (Rpd3SL) complex. KDM5B removes intragenic H3K4me3, represses cryptic initiation, and safeguards productive mRNA elongation.

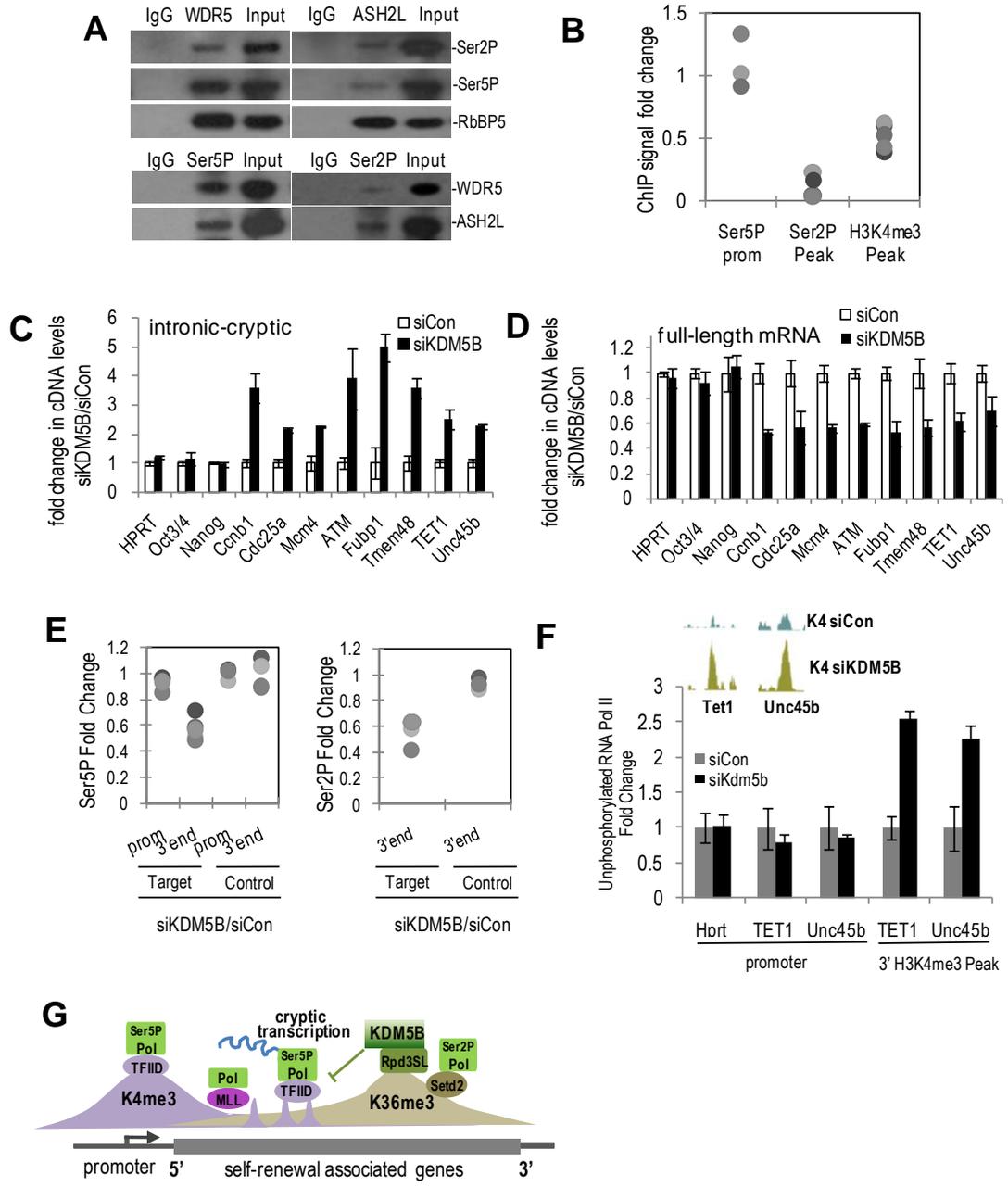
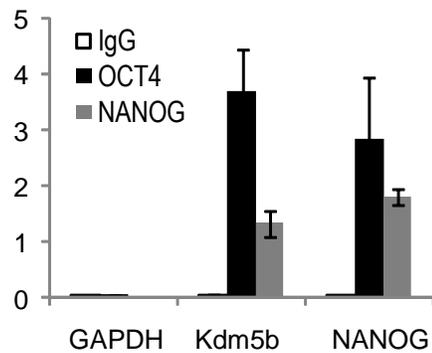


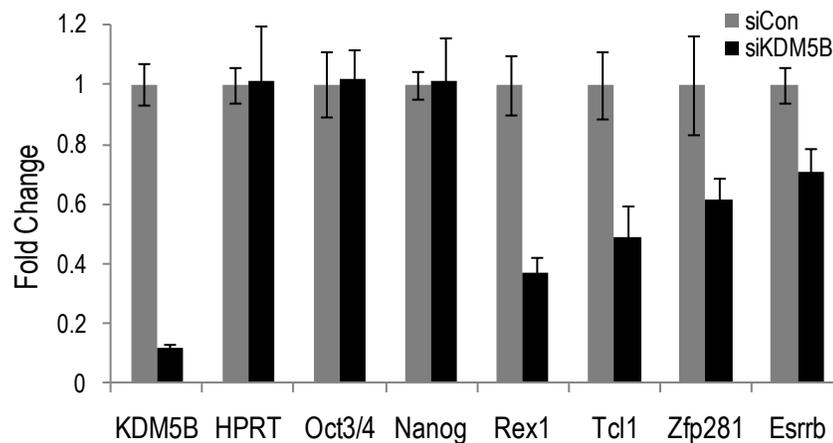
Figure-6

Supplemental Figure and Figure legends

(A)

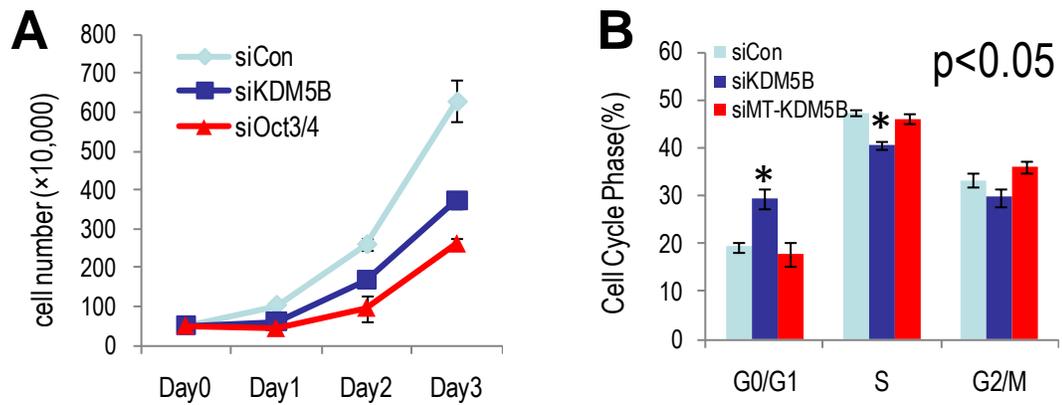


(B)

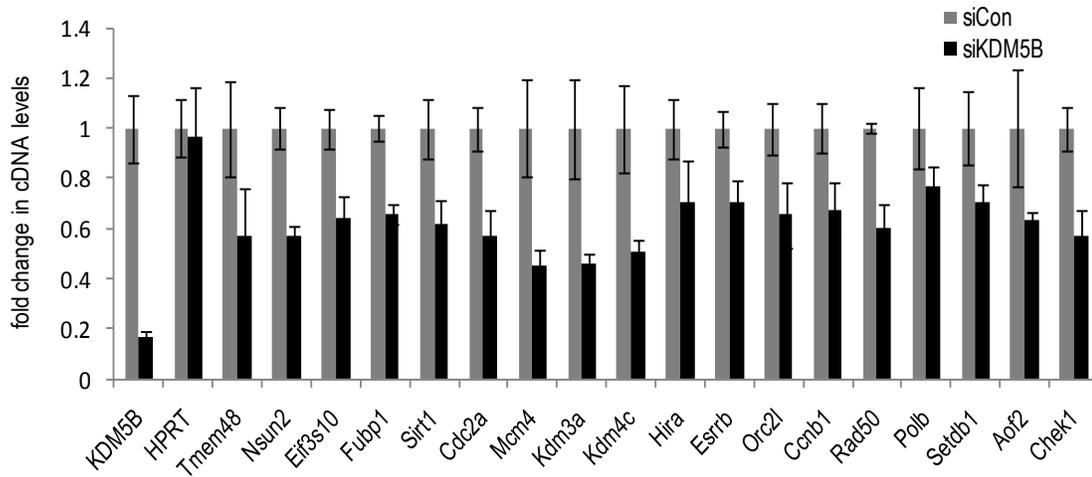


Supplemental Figure 2.1 (A) ChIP signal for OCT4 and NANOG at the human KDM5B genomic locus in H1 human ESCs. The NANOG promoter region is a positive control.

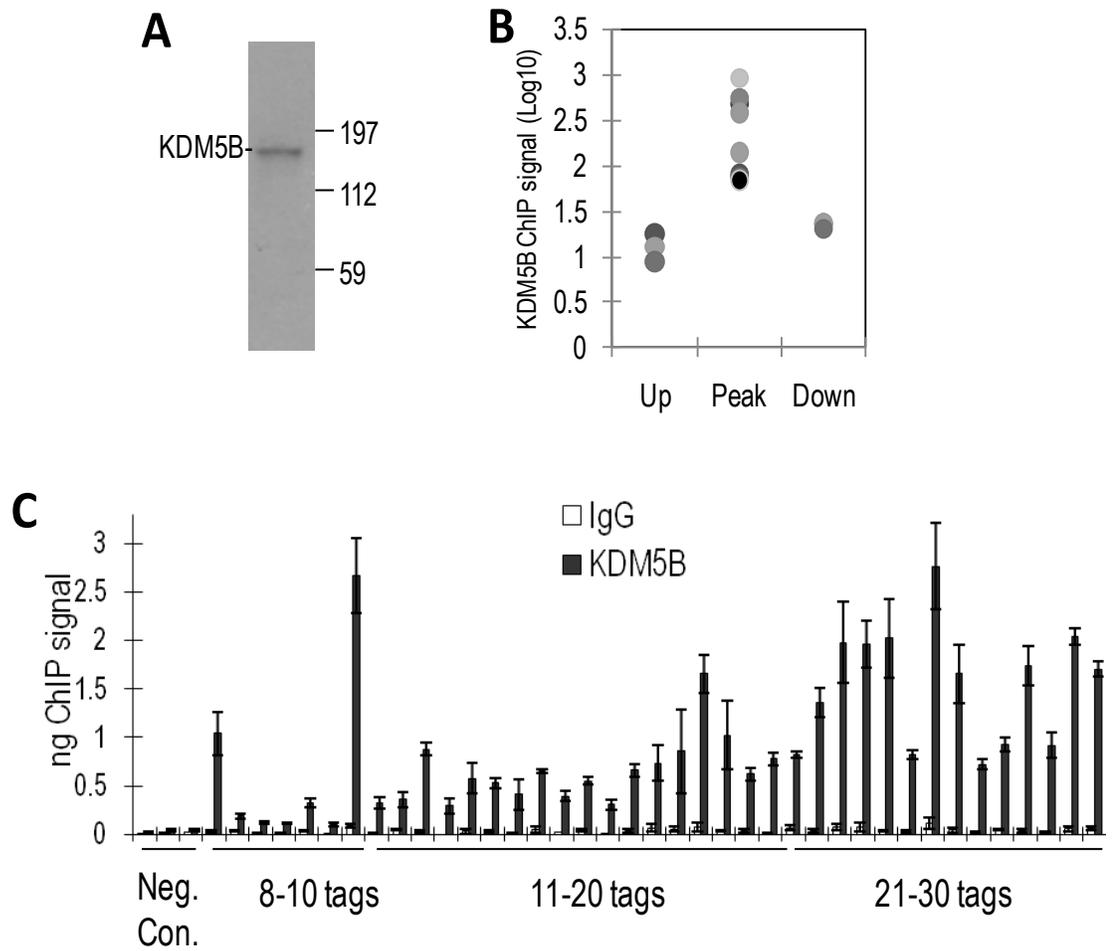
(B) Real-time RT-PCR analysis of pluripotency-associated gene expression following knockdown of *Kdm5b* in ESCs.



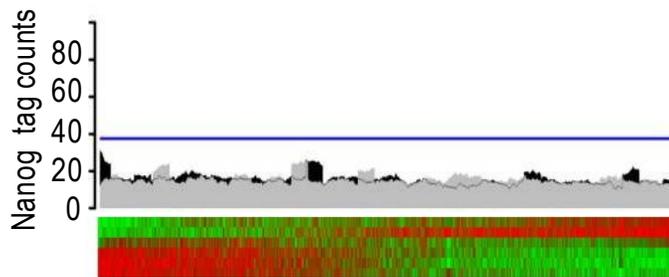
Supplemental Figure 2.2 (A) J1 ESCs were transfected with control, Kdm5b or Oct4 siRNA and cell number was counted at the indicated time points. **(B)** FACS analysis of cell cycle phase from J1 ESCs transfected with control, Kdm5b siRNA or Kdm5b siRNA mutant (siMT-Kdm5b) which didn't abolish endogenous Kdm5b.



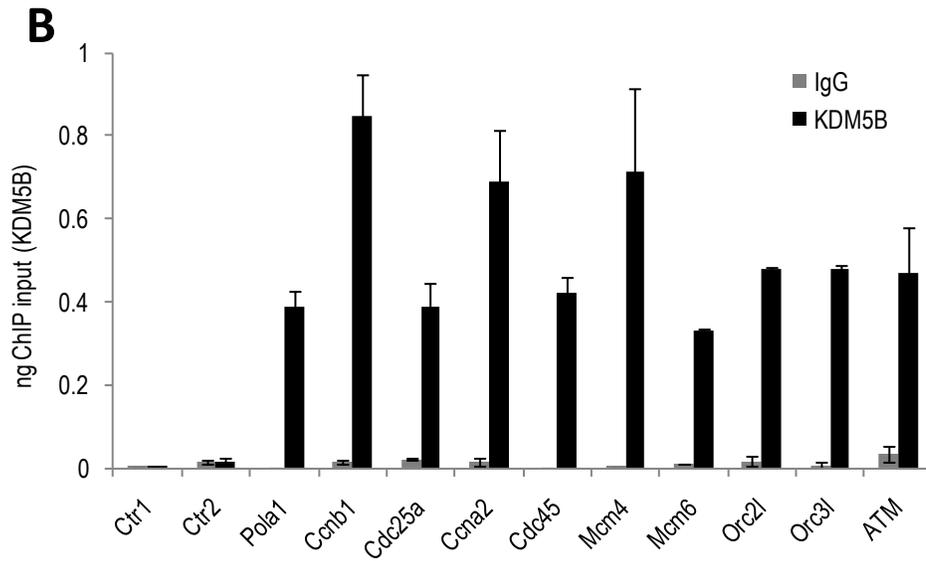
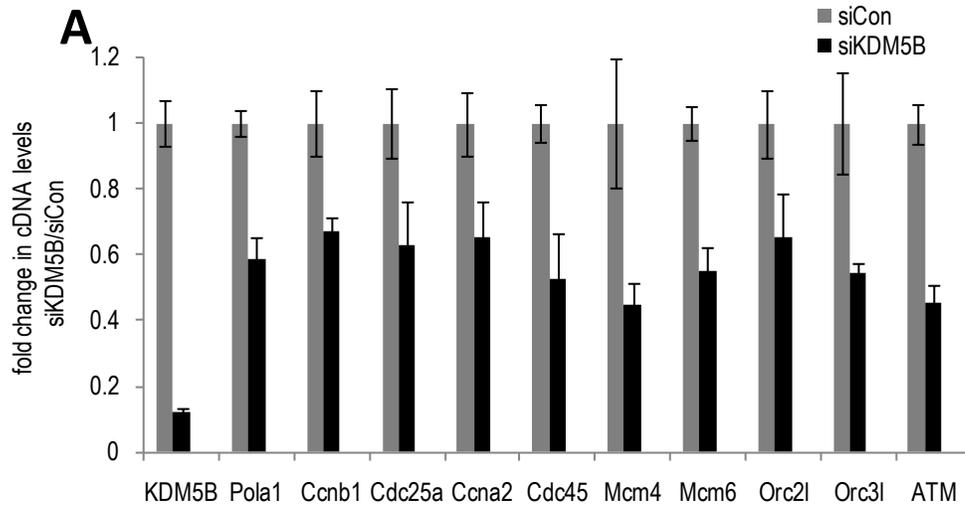
Supplemental Figure 2.3 RT-PCR analysis of significantly-regulated genes from siKdm5b Affymetrix microarray. J1 ESCs were transfected with control or Kdm5b siRNAs (siKdm5b).



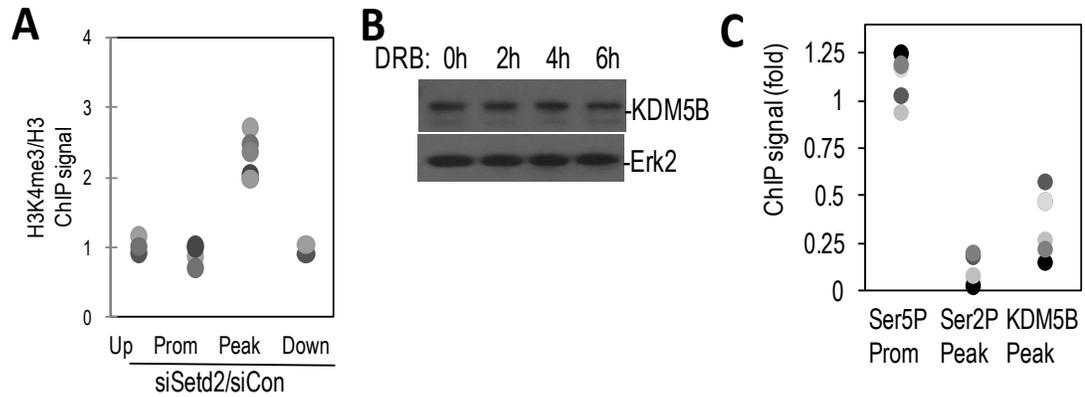
Supplemental Figure 2.4 (A) Immunoblot analysis of ESC nuclear extracts for KDM5B. **(B)** ChIP-PCR analysis using a KDM5B antibody to a different epitope (Santa Cruz sc-67035) at negative control regions (Up and Down) and ChIP-Seq peaks (Peak). Each dot represents a distinct ChIP-Seq locus (pairwise comparisons to Peak data $p < 0.01$). **(C)** ChIP-PCR analysis of a random subset of KDM5B ChIP-Seq peaks. Tag counts denote peak amplitude.



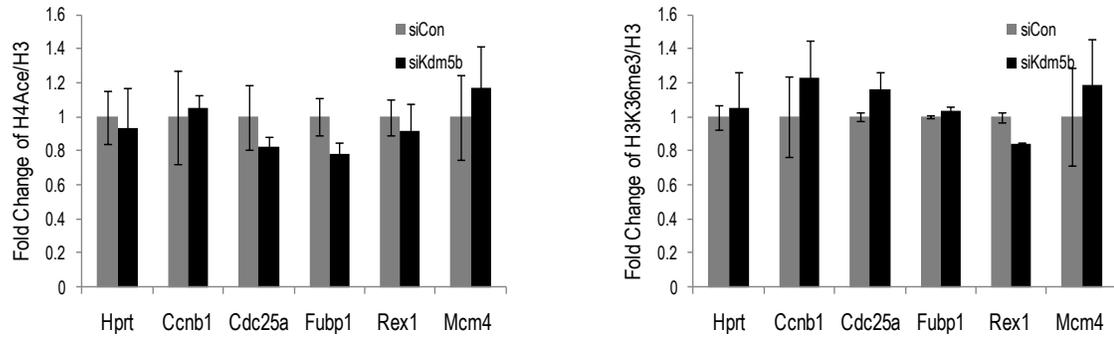
Supplemental Figure 2.5 Histogram of Nanog CHIP-Seq peak tag density relative to microarray data following *Kdm5b* knockdown. The black profile above the heat map depicts average number of CHIP-Seq KDM5B peak tags found per gene. The gray profile represents a single randomized control and the blue line delineates an FDR of $p < 0.01$ (permutation statistic).



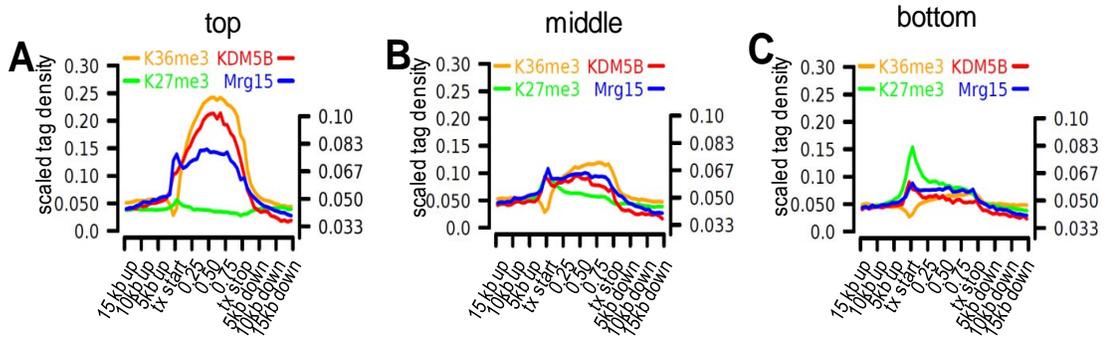
Supplemental Figure 2.6 (A) RT-PCR analysis of selected KDM5B target genes. **(B)** KDM5B ChIP-qPCR analysis of selected Kdm5b target genes at intragenic KDM5B ChIP-Seq peak regions.



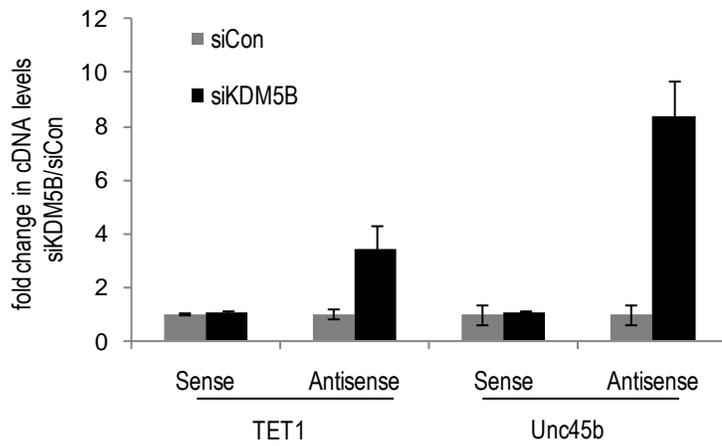
Supplemental Figure 2.7 (A) H3K4me3 ChIP-PCR signal at negative control regions (Up, Prom, Down) and Kdm5b peaks following knockdown of Setd2 (siSetd2). Data are expressed as the ratio of H3K4me3 ChIP signal of siSetd2 over siCon. Each dot represents a distinct locus. **(B)** ESCs were treated with 50 μ M DRB for 2-6 hours. Kdm5b total protein level was examined by immunoblot. **(C)** ChIP-PCR analysis of Ser2P, Ser5P, and KDM5B at KDM5B ChIP-Seq peaks and associated promoters after 6hr DRB treatment. Data is expressed as fold change versus untreated (0 hr). Each dot represents a distinct ChIP-Seq locus.



Supplemental Figure 2.8 ChIP analysis of H4Ace (A) or H3K36me3 at KDM5B peaks at the 3' end of target genes following Kdm5b siRNA treatment in J1 ESCs. Primers towards 3'end of Hprt gene body was used as a control.



Supplemental Figure 2.9 (A-C) ChIP-Seq tag density relative to the top (A), middle (B), and bottom third (C) of RefSeq genes expressed in J1 ESCs (Affymetrix microarray data). All Ref-Seq gene lengths were scaled to 1. The left Y-axis corresponds to H3K36me3 and H3K27me3 ChIP-Seq data while the right Y-axis corresponds to KDM5B and MRG15 ChIP-Seq data.



Supplemental Figure 2.10 J1 ESCs were transfected with control or *Kdm5b* siRNA and reverse transcription was performed using strand-specific primers towards TET1 or Unc45b intronic region.

Methods

Plasmids

Mouse *Kdm5b* was amplified using PCR from murine cDNA and cloned into pUB6-V5 (Invitrogen). HA-tagged MRG15 vector is from Dr. Kaoru Tominaga and Dr. Olivia M. Pereira-Smith. All vectors were verified by sequencing.

Cell culture and reagents

J1 and ZHBTc4 mouse ESCs were cultured at 37°C with 5% CO₂. All cells were maintained on gelatin-coated dishes in Dulbecco's modified Eagle medium (Invitrogen), supplemented with 10% ESC-tested fetal bovine serum (PAA), 0.1 mM MEM NEAA (Invitrogen), 2 mM L-glutamine (Invitrogen), 100U/ml/100ug/ml Pen Strep (Invitrogen), 0.055 mM β-mercaptoethanol (Invitrogen), and 1000 U/mL LIF (Chemicon). Human ESCs (H1) were cultured on matrigel-coated plates in mTeSR1 medium (StemCell Technologies) and passaged using collagenase IV according to the manufacturer's protocol (StemCell Technologies). 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB, 50 μM; Sigma) was directly applied to ESC medium for indicated periods of time. AP activity was monitored using the commercial alkaline phosphatase Detection Kit (Millipore) according to the manufacturer's instructions.

Antibodies

Antibodies used for immunoblotting include: KDM5B (abcam 50958 or Santa Cruz sc-67035), Oct4 (Santa Cruz sc-5279), Nanog (Chemicon AB5731), Histone 3 (Abcam

ab1791) H3K4me3 (Active Motif 39159), H3K9me3 (Abcam ab8898-100), H3K27me3 (Abcam ab6002-100), H3K36me3 (Abcam ab9050-100), MRG15 (gift from Dr. Kaoru Tominaga and Dr. Olivia M. Pereira-Smith), RbBP5 (Bethyl A300-109A), WDR5 (Abcam, ab22512-100), Ash2L (Bethyl A300-112A). Antibodies utilized for ChIP or IP: Kdm5b (Abcam ab50958 and Santa Cruz sc-67035), Nanog (Chemicon, AB5731), Oct4 (Santa Cruz, sc-5279), Histone 3 (Abcam, ab1791), H3K4me3 (Active Motif, 39159), H3K36me3 (Abcam, ab9050-100), MRG15 (AVIVA Systems Biology, ARP32832-T100), Ser2P RNAP (abcam ab5095-100 and Covance H5), Ser5P RNAP (Covance H14), WDR5 (Abcam, ab22512-100), Ash2L (Bethyl A300-112A),

RNA interference

RNA-interference of J1 ESCs was conducted as previously described (Loh et al, 2006). The following 19 mer sequences were cloned into the pSUPER.puro backbone according to the manufacturer's instructions (Oligoengine). *Kdm5b*-1: 5'-TCTTTGCCCTCGGTGTGAC-3'; *Kdm5b*-2: 5'-GAGATGCACTCCGATACAT-3'; *Mrg15*-1: 5'- GAGTACCATCGGAAAGCCG-3'; *Mrg15*-2: 5'- ACAATATGCAGAGGGCAAG-3'; *Setd2*-1: 5'- GTCTTAAGCTCATTTCAGAA-3', *Setd2*-2: 5'- GGGAGTGTCTGATGTTGAA-3'. 19-mer sequences for GFP, Oct4, and Nanog were previously described (Loh et al, 2006). Puromycin (Sigma) selection was introduced 1 d after transfection at 1 µg /mL, and maintained for 2-3 days prior to harvesting.

RNA isolation, reverse transcription and quantitative real-time PCR

Total or nuclear RNA was isolated using the Trizol (Invitrogen) method following manufacture's protocol and reverse-transcribed using MMLV (Invitrogen). Real time PCR was conducted as described previously (Impey et al, 2004). All RT-PCR data utilized standard curve real-time PCR. Primers will be provided upon request.

Proliferation assay and Cell Cycle Analysis

Equal numbers of J1 ESCs (500,000) were plated in 6-well plates and transfected with the indicated siRNAs. Three wells were counted with a hemocytometer at the indicated times. For FACS analyses, ESCs were trypsinized, washed with PBS, fixed with 70% ice-cold ethanol, and stained with propidium iodide (10 µg/ml, Sigma). DNA cell cycle analysis was conducted using a FACS Calibur instrument (Becton Dickinson) in the Oregon Stem Cell Center Flow Cytometry Core. Cell cycle compartments were deconvoluted from single-parameter DNA histograms of 20,000 cells and cell cycle data was analyzed using ModFit (Verity Software House).

Chromatin Immunoprecipitation assay

ChIP was conducted as described previously (Impey et al, 2004). Briefly, ESCs were fixed with 1% formaldehyde for 10 min at room temperature. Chromatin was sonicated and immunoprecipitated using indicated antibodies overnight. All ChIP-PCR analysis was conducted using real-time PCR. Primers will be provided upon request.

Microarray analysis

Control and Kdm5b siRNA (designed from Darmacon) treated samples were run on MOE430 2.0 Affymetrix chips using standard procedures. ESC retinoic acid time course microarray data were obtained from GEO (GSE2972). CEL file data was processed using RMA and significant genes were isolated using the limma (heatmap analysis) or SAM (gene lists) R packages.

ChIP-Seq and RNA-Seq analysis

For RNA-Seq ~10 ng of Poly(A) RNA was converted to double stranded cDNA as described previously (Nagalakshmi et al, 2008). ChIP and cDNA samples were sequenced using standard Solexa protocols. 25 bp reads were mapped to the mouse genome (UCSC mm9) and unique tags were selected for analysis. For ChIP-Seq, areas of enrichment at an FDR of 0.05 were determined using a custom sliding-window approach based on previous work (Fejes et al, 2008) but implemented in C++. For RNA-Seq analyses, counts of mapped tags in RefSeq exons were determined and significantly-regulated genes were selected using an FDR-adjusted chi-square statistic. Statistical analyses were carried out in the R programming environment. H3K36me3 and H3K36me27 ChIP-Seq data used in this study were generated from Solexa/Illumina output from a previous study (Mikkelsen et al, 2007)

Bioinformatic and Statistical Analyses

Microarray heatmaps were generated by selecting significantly-regulated genes at an FDR of 0.1 using the Bioconductor limma package and sorting based on fold-change (Gentleman et al, 2004). The RNA-Seq heatmap was generated in R by selecting genes

with an FDR < 0.001 and a fold-change >2 and sorting based on the FDR-adjusted chi-square p-value. All heatmap data was log-scaled and row-scaled. Pie charts were generated using custom annotation scripts and statistical significance was assessed with the Wilcoxon rank-sum test. Profile plots of ChIP-Seq data relative to RefSeq genes were generated in R by setting all gene lengths equal to 1 and plotting tag density. Profile plots of histone methylation relative to Ref-Seq genes sorted by KDM5B Peak area were generated by selecting equal sets of histone methylation ChIP-Seq peaks and plotting peaks counts that either overlap with a Ref-Seq gene or are within 1kb of a gene end. ChIP-Seq profile plots relative to heatmaps were generated by displaying averaged ChIP-Seq peak counts that overlap with at least 1 bp of a Ref-Seq genes present in the heatmap (black profile). A “background” for ChIP-Seq profile plots was generated by randomizing the position of ChIP-Seq peaks and the blue significance line was generated using a permutation test. The UCSC genome browser was used to visualize RNA-Seq and ChIP-Seq data. Gene ontology analysis was conducted using DAVID (Dennis et al, 2003). Comparison of low-throughput biological data utilized the Student’s t-test or ANOVA followed by a post-test where appropriate. The Storey Q-value was used for all FDR adjustments (Storey & Tibshirani, 2003). ChIP-Seq, RNA-Seq and microarray data have been deposited in GEO (accession number will be added).

CHAPTER THREE

KDM5B, an Epigenetic Module Regulating Embryonic Stem Cell Pluripotency

KDM5B, an Epigenetic Module Regulating Embryonic Stem Cell Pluripotency

Oregon Stem Cell Center, Papé Family Pediatric Research Institute, Cell and Developmental Biology, Oregon Health and Science University.

*Correspondence: Soren Impey, Oregon Stem Cell Center, Oregon Health and Science University, 3181 SW Sam Jackson Park Rd./ L321, Portland OR 97239-3098, impeys@ohsu.edu, (Phone:503-494-7540, Fax:503-494-5044)

Abstract

Nanog is believed to orchestrate an epigenetic program critical for the establishment and maintenance of embryonic stem cell (ESC) pluripotency. Nevertheless, the molecular mechanisms by which Nanog target genes regulate epigenetic programming remain obscure. We utilized RNAi knockdown and functional overexpression assays to identify the histone H3 lysine 4 demethylase KDM5B as a chromatin modulator important for ESC identity. Ectopic expression of *Kdm5b* is sufficient to support ESC pluripotency in the absence of LIF signaling for over two months and preserves the capacity for teratoma formation. Epistasis experiments demonstrate that overexpression of *Kdm5b* can compensate for loss of Nanog. We utilized KDM5B ChIP-Seq and *Kdm5b* siRNA RNA-Seq to identify KDM5B regulated ESC self-renewal associated gene network. We find that cell cycle and DNA replication genes *Mcm7* and *Orc1* contribute to KDM5B-mediated ESC self-renewal. Moreover, proteomics analysis reveals that KDM5B interacts with the DNA ore-replication complex We propose that KDM5B is a major mediator of ESC self-renewal and functions by maintaining a high rate of mitosis that contributes to the naïve epigenetic state.

Introduction

Self-renewing ESCs derived from the inner cell mass (ICM) of pre-implantation embryos have the capacity to give rise to all three germ lineages (Orkin & Hochedlinger, 2011). Understanding the genetic and epigenetic programs governing ESC identity is important for their use in regenerative medicine. The ESC epigenome has an open, hyperdynamic chromatin structure (Niwa, 2007b; Orkin & Hochedlinger, 2011) and accumulates overlapping active H3K4me3 and repressive H3K27me3 (Bernstein et al, 2007). The unique ESC chromatin architecture not only facilitates core pluripotency factors to access and execute the genome but also regulates ESCs cell fate (Orkin & Hochedlinger, 2011). Moreover, chromatin modulators have been shown to facilitate or inhibit acquisition of pluripotency in somatic reprogramming (Onder et al, 2012; Singhal et al, 2010). These studies strongly suggest that reprogramming of the epigenome by chromatin modulators is critical for the induction of the pluripotent state..

Forced expression of core pluripotency transcription factor Nanog can support indefinite ESC self-renewal in the absence of LIF signaling (Chambers et al, 2003; Mitsui et al, 2003) *Nanog* null embryos show peri-implantation lethality and cannot mature into epiblasts, the founder tissue of somatic lineages (Mitsui et al, 2003). Silva and colleagues show that Nanog deletion failed to reprogram somatic cells into the ground pluripotent state (Silva et al, 2009). These studies suggest an indispensable role for Nanog in both embryonic and induced pluripotency (Silva et al, 2009). Moreover, cell fusion-mediated reprogramming is promoted by and dependent on Nanog (Silva et al, 2006; Silva et al, 2009). During Nanog induced fusion-mediated reprogramming, extensive remodelling of the somatic epigenome occurs, including DNA demethylation, X inactive chromosome reactivation, re-setting of the bivalent modifications (Hochedlinger

& Plath, 2009). This has led to the proposal that Nanog coordinates a genetic and epigenetic program that induces naïve pluripotency. Although previous studies have reported genome-wide Nanog occupancy (Chen et al, 2008; Loh et al, 2006), the role of Nanog target genes that regulate chromatin function has not been well studied.

We identified KDM5B as a regulator of ESC self-renewal via a functional expression screen of Nanog target genes. *KDM5B* is also markedly upregulated in multiple human cancers and has been linked to cancer cell proliferation and self-renewal (Hayami et al, 2010; Lu et al, 1999; Roesch et al, 2010; Xiang et al, 2007; Yamane et al, 2007). *Kdm5b* null embryos display peri-implantation lethality and *Kdm5b* null ESCs cannot be derived (Catchpole et al, 2011), suggesting *Kdm5b* plays a critical role in early embryonic development and pluripotent stem cell specification. Acute depletion of *Kdm5b* by siRNA compromises the ability of ESCs to proliferate and self-renew (Xie et al, 2011). Interestingly ectopic *Kdm5b* expression can increase H3Ser10 phosphorylation in ESCs and reduce neural progenitor cell differentiation (Dey et al, 2008). Here, we show that forced expression of *Kdm5b* can mimic Nanog to sustain ESC pluripotency in the absence of LIF for over 20 passages. Moreover, we show that ectopic *Kdm5b* expression can bypass the requirement for Nanog to support ESC maintenance. Forced expression of *Kdm5b* also enhanced reprogramming efficiency. Bioinformatic analyses of ChIP-Seq and RNA-Seq data suggested KDM5B promotes self-renewal by directly regulating genes involved in DNA replication. Interestingly, KDM5B interacts with many of the same regulators of DNA replication that it targets suggesting a conserved positive feedback loop that mediates self-renewal. We propose that KDM5B is an important epigenetic regulator of ESC self-renewal and represents a major effector of the Nanog self-renewal phenotype.

Results

Functional expression assay identifies novel Nanog-regulated chromatin modulator genes in ESC maintenance

Nanog was discovered by virtue of its ability to support ESC self-renewal without the LIF signaling but downstream targets that mediate this phenotype have not been well-characterized (Chambers et al, 2003). We performed a bioinformatic analysis of Nanog ChIP-Seq and *Nanog* knockdown RNA-Seq libraries to identify genes that are both targeted and transcriptionally-regulated by Nanog (Figure 3.1A). Gene ontology analysis identified 26 putative Nanog target genes that regulate chromatin structure or modification (Figure 3.1B).

We next tested whether forced expression of these 26 candidate genes could mimic Nanog's ability to support ESC maintenance in the absence of LIF. We generated ESC lines stably expressing each of the 24 (out of 26) chromatin regulatory genes (Supplemental Figure S3.1A). Under normal growth conditions, all clones show normal ESC morphology and strong staining for alkaline phosphatase (AP), a marker of the undifferentiated state. 6 days after plating these cell lines at low density in the absence of LIF, only the *Myst4*, *Phf15*, *Kdm5b* and *Wdr82*-expressing cell lines still contained detectable AP-positive colonies while the other lines had a flattened differentiated morphology and undetectable AP activity (Figure 3.1C and 3.1D, Supplemental Figure S3.1B).

RNAi assay identifies critical chromatin regulators in ESC maintenance

We next tested whether these candidate self-renewal genes were required for ESC self-renewal using RNAi. Knockdown of 6 chromatin regulators (*Myst4*, *Phf15*, *Kdm5b*, *Wdr82*, *Setd5* and *Arid5b*) that have not been linked to self-renewal triggered a loss of AP staining and/or a deficit in proliferation (Figure 3.1.E and Supplemental Figure S3.1D). Another six candidates (*Tet1*, *Wdr5*, *Arid1b*, *Kdm3a*, *Rcor2*, *Phc1*) also showed self-renewal phenotypes and were previously linked to ESC self-renewal (Ang et al, 2011; Freudenberg et al, 2012; Ito et al, 2010; Loh et al, 2007; Walker et al, 2007; Yan et al, 2008; Yang et al, 2011). Importantly, we showed that all RNAi knockdowns decreased target transcription (Supplemental Figure S3.1C). Importantly we confirmed that depletion of *Myst4*, *Phf15*, *Wdr82* decreased expression of pluripotency-associated genes with a concomitant increased expression of differentiation-associated markers (Supplemental Figure S3.1E). Both loss- and gain-of-function assays converge on an overlapping set of novel chromatin regulator genes. In this study, we focused on *Kdm5b* because *Kdm5b* is one of the best Nanog ChIP-Seq targets and has been implicated in the proliferation and self-renewal of a variety of human cancers (Hayami et al, 2010; Lu et al, 1999; Roesch et al, 2010; Xiang et al, 2007; Yamane et al, 2007).

Expression of *Kdm5b* in pluripotency associated cells and tissues

We next tested whether *Kdm5b* expression was restricted to pluripotent cells and tissues like its upstream regulator Nanog. In pre-implantation mouse embryos, *Kdm5b* transcript becomes detectable at the 2-cell stage with peak expression at the late morula and blastocyst stages (Figure 3.2A and Supplemental Figure S3.2A).

Immunofluorescence analysis showed that KDM5B protein is confined to the ICM of mouse blastocysts, where pluripotent stem cells are derived (Figure 3.2B). Expression of *Kdm5b* is highly abundant in pluripotent ESCs and embryonic carcinoma cells but not in somatic cells or tissues, except for the testis, a reservoir of pluripotent cells (Kee et al, 2010) (Figure 3.2C). Retinoic acid-induced differentiation of mouse and human ESCs results in the rapid loss of *Kdm5b* expression, but not other KDM5 family members (Figure 3.2D and Supplemental Figure S3.2B and 3.2C).

Among the four mammalian Kdm5 family members, *Kdm5b* is most enriched in ESCs whereas *Kdm5d* expression is not detectable. Knockdown of *Kdm5a* or *Kdm5c* did not affect ESC proliferation or pluripotency (Supplemental Figure S3.2E and S3.2F). Nanog knockdown reduced *Kdm5b* expression but had no effect on others (Figure 3.2E). *Kdm5b* knockdown also did not alter the expression of other KDM5 family members (Supplemental Figure S3.2D). Thus, *Kdm5b*'s expression pattern suggests a role in pluripotent stem cells.

Overexpression of *Kdm5b* supports undifferentiated ESC maintenance independently of LIF signaling

Our functional expression assay revealed that *Kdm5b* overexpression appeared to reduce the requirement for LIF for ESC undifferentiated self-renewal. Quantification of AP positive colonies derived from ESCs overexpressing *Kdm5b* without LIF for 5 days yielded approximately the same number of undifferentiated colonies as Nanog (Figure 3.3A). Moreover, forced expression of *Kdm5b* increased ESC proliferation and the percentage of cells in the S phase (Supplemental Figure S3.3A-C). Similarly,

overexpression of wild type but not a catalytic-inactive mutant *Kdm5b*, maintained proliferation to the same degree as forced expression of Nanog in the absence of LIF (Figure 3.3B). Ectopic expression of *Kdm5b* prevented down-regulation of pluripotency-associated genes to the same degree as forced-expression of Nanog (Figure 3.3C). Secondary ESC replating assays confirmed that *Kdm5b* overexpressing ESCs could still form undifferentiated colonies after 6 days (Figure 3.3D). Forced expression of *Kdm5b* was also able to inhibit ESC differentiation even when induced by stronger differentiation paradigms (Supplemental Figure S3.3D-G). Taken together, these data suggest that ectopic expression of *Kdm5b* promotes the undifferentiated pluripotent state even in the presence of differentiation signals.

KDM5B promotes indefinite self-renewal of undifferentiated ESCs in the absence of LIF

To address whether *Kdm5b* can support ESC pluripotency following long-term LIF withdrawal, we generated stable cell lines that express *Kdm5b* (termed E5B). Similar to Nanog stable cell lines, E5B cells grown in the absence of LIF for 16 passages (~8 weeks) continue to show strong AP staining (Figure 3.4A). Moreover, maintenance of the undifferentiated state depends on the *Kdm5b* catalytic activity (Supplemental Figure S3.4B). Cytogenetic analysis demonstrates that these E5B cells are karyotypically normal (Figure 3.4C). Cre-recombinase mediated excision of *Kdm5b* in E5B cells (E5BC cells) triggered expression of a DsRed transgene and caused E5BC cells to rapidly differentiate in the absence of LIF (Supplemental Figure S3.4A). Importantly, differentiation of E5BC cells can be blocked by addition of LIF (Fig 3.4A), suggesting that sustained expression of *Kdm5b* is able to preserve endogenous ESC self-renewal

pathways. Analysis of pluripotency marker genes in E5B cells grown in the absence of LIF confirmed that an undifferentiated gene profile while Cre-mediated excision of *Kdm5b* triggered a marked reduction of pluripotency markers (Figure 3.4B). These data strongly suggest that ectopic expression of *Kdm5b* promotes cell autonomous ESC self-renewal.

***Kdm5b* stable cells support teratoma formation**

A hallmark of ESCs is their ability to differentiate into all three somatic lineages. To examine whether E5B cells are pluripotent following long-term culture without LIF, we used Cre recombinase to delete exogenous *Kdm5b*, cultured briefly in LIF medium and injected these cells into immunodeficient mice for teratoma formation assay. 3 weeks after injection, E5BC cells formed tumors of as the same size as parental ESCs (Supplemental Figure S3.4C). Hematoxylin and eosin staining of E5BC cell-derived tumors revealed representative sections of all three somatic layers (Figure 3.4D). These data show that *Kdm5b* expression can maintain ESCs pluripotency even after prolonged culture in the absence of LIF.

***Kdm5b* compensates for loss of Nanog and is required for Nanog-mediated self-renewal**

Because *Kdm5b* can support LIF-independent self-renewal similar to *Nanog*, we examined the epistatic relationship between *Nanog* and *Kdm5b*. If *Kdm5b* functions downstream of *Nanog*, forced expression of *Kdm5b* would be expected to rescue the *Nanog* knockdown phenotype. ESCs stably expressing wild-type *Kdm5b*, but not

catalytically inactive *Kdm5b*, attenuated differentiation-mediated loss of AP activity following *Nanog* depletion (Figure 3.5A). This suggests that *Kdm5b* can compensate for loss of *Nanog*. Gene expression analyses confirmed that *Kdm5b* expression attenuated induction of differentiation markers following *Nanog* knockdown (Figure 3.5B). Consistent with the idea that *Nanog* is upstream of *Kdm5b*, constitutive *Nanog* expression significantly increased *Kdm5b* levels in both differentiating ESC and undifferentiated cells (Figure 3.3C and Supplemental Figure S3.5B). We next asked whether *Kdm5b* function was required for Nanog-mediated self-renewal. Knockdown of *Kdm5b* impaired ESC self-renewal in cells that constitutively express *Nanog* (Figure S3.5C). Importantly, stable *Nanog* expression in the same cell line blocked differentiation triggered by LIF withdrawal and retinoic acid treatment (Figure 3.3A-C). *Kdm5b* depletion triggered similar increases in lineage-associated markers in wild-type cells and *Nanog* expressing ESCs (Figure 3.5D). Our data suggests that *Kdm5b* is downstream of *Nanog* and is required for *Nanog*-mediated self-renewal.

Niwa et al. proposed that LIF signaling regulates ESC self-renewal through parallel signaling pathways: a *Klf4*-*Sox2* pathway and a *Tbx3*-*Nanog* pathway (Niwa et al, 2009). We found no induction of *Kdm5b* after *Klf4* overexpression (Supplemental Figure S3.5B). Moreover, knockdown of *Klfs* (Jiang et al, 2008) can readily differentiate ESCs stably expressing *Kdm5b*, as evidenced by loss of ESC morphology and induction of lineage-associated genes expression (Supplemental Figure S3.5E-F). Conversely, *Kdm5b* knockdown can still produce a proliferation defect even when *Klf4* is overexpressed (Supplemental Figure S3.5C-D). This data suggests that *Klf4* and *Kdm5b* represent parallel pathways.

***Kdm5b* enhances reprogramming of somatic cells to pluripotency**

Forced expression of *Nanog* enhances the induction of pluripotency and epigenetic reprogramming (Silva et al, 2006; Silva et al, 2009; Takahashi et al, 2007). Our data suggests that *Kdm5b* is an important downstream effector of *Nanog* and raises the question whether *Kdm5b* can contribute to the induced pluripotency.

Because retrovirus vectors do not express large proteins well (~170KD for full length KDM5B protein), we utilized a retrovirus vector expressing a shorter C-terminal truncated *Kdm5b* which preserves the intact demethylase activity (Yamane et al, 2007). Strikingly, we found that *Kdm5b* could significantly enhance AP positive ESC-like colonies formation in conjugation with *Oct4* alone from neural stem cells (Kim et al, 2008b). In *Oct4/Klf4/Kdm5b* condition, reprogramming efficiency was greatly boosted (Figure 3.5E). Ectopic expression of catalytic inactive *Kdm5b* mutant did not enhance reprogramming efficiency, suggesting that the H3K4me3/me2 demethylase activity is required for induced reprogramming. The gene expression signatures of these iPSC lines and their ability to form teratomas are being assessed.

KDM5B does not regulate upstream self-renewal signaling

We next examined the molecular mechanisms by which KDM5B supports ESC pluripotency. First, we examined whether KDM5B modulates the major growth factor-mediated signaling pathways that support ESC pluripotency, namely, the LIF-Stat3 and BMP4-Smad-I δ pathway (Niwa et al, 1998; Ying et al, 2003). *Kdm5b* overexpression does not alter the kinetics of phosphorylated Stat3 activation or I δ gene induction (Figure

3.6A-B). These data are consistent with the idea that KDM5B functions as an epigenetic effector rather than via feedback regulation on intracellular signaling pathways.

KDM5B directly upregulates DNA replication and pre-replication machinery

We previously showed that KDM5B is a positive regulator of transcriptional elongation and targets to a subset of genes involved in self-renewal (Xie et al, 2011). Interestingly, most KDM5B-regulated genes are not targets of core pluripotency-associated transcription factors, suggesting that KDM5B represents a distinct epigenetic node downstream of Nanog. By analyzing KDM5B ChIP-Seq and newly generated Kdm5b knockdown RNA-Seq libraries, we obtained a list of KDM5B activated self-renewal associated genes (Figure 3.6C). Interestingly, many of the KDM5B targets are associated with cell cycle control and DNA synthesis, including *Pola1*, *Cdc45*, *Mcm3-7*, *Orc1/2/3/5/6*, *Cdc25a*, *Ccnb1* etc (Xie et al, 2011). *Mcm4* and *Mcm6* contribute to the rapid proliferation in the early embryo (Coue et al, 1996) and both *Mcm4* and *Cdc25a* null embryos fail to develop the inner cell mass (Lee et al, 2009a; Shima et al, 2007). Many KDM5B target genes, including *Kdm4c*, *Kdm3a*, and *Smc1a* have recently been implicated in ESC self-renewal (Hu et al, 2009; Ito et al, 2010; Loh et al, 2007).

To examine whether these KDM5B target genes contribute to KDM5B-mediated ESC proliferation and self-renewal, we performed gain and loss of function experiments. Consistent with an earlier report (Fazio & Panning, 2010), siRNA knockdown of *Orc1* or *Mcm7* reduced the proliferation rate of ESCs but not MEFs (Figure 3.6D), suggesting that these genes regulate ESC-specific self-renewal. Interestingly, overexpression of *Orc1* or *Mcm7* can partially rescue the proliferation defect produced by *Kdm5b*

knockdown (Figure 3.6E). On the other hand, knockdown of these DNA replication machinery genes markedly decreased ESC proliferation even when *Kdm5b* is overexpressed (Figure 3.6D). These data suggest that cell cycle control/DNA synthesis machinery genes contribute to KDM5B-mediated self-renewal.

Characterization of the KDM5B demethylase complex in ESCs

To further elucidate KDM5B's mechanism in regulating ESC identity, we affinity purified KDM5B-associated complexes using an antibody that recognizes endogenous KDM5B (Figure 3.7A) followed by mass-spectrometry sequencing. Flag antibody-mediated affinity purification of 3XFlag-KDM5B complexes in ESCs generates large amount of overlap between the two data sets. Many of the polypeptides including NCOR1, HDAC6, LSD1, CHD4 and MTA1 were reported by other labs that interact with KDM5B (Barrett et al, 2007; Li et al, 2011). The predicted KDM5B-interacting proteins showed significant enrichment for regulators of mRNA processing, ATP dependent chromatin remodeling complexes/histone modifications, and regulators of cell cycle progression/DNA synthesis (Figure 3.7B and Appendix Figure 2). Moreover, a significant fraction of these targets are direct transcriptional targets of KDM5B (e.g. *Mcm5/7*, *Orc1*, *Chd4* etc.), suggesting a positive-feedback regulatory loop to reinforce KDM5B's regulatory role in cell cycle/DNA synthesis control.

Consistent with the observation that KDM5B highly correlates with H3K36me3, KDM5B proteomic analysis pulled down the H3K36 HMT protein NSD3 (Vermeulen et al, 2010; Vermeulen et al, 2007). NSD3 can specifically recognize H3K36me3 via its PWWP domain and has been linked to transcription elongation (Fang et al, 2010; Lucio-

Eterovic et al, ; Rahman et al, 2011; Rayasam et al, 2003; Wang et al, 2007). Reciprocal Co-IP confirmed the interaction between of KDM5B and NSD3 (Appendix Figure 3B). Furthermore, *Nsd3* shRNA knockdown reduces KDM5B occupancy at intragenic regions of KDM5B target genes and decreases KDM5B target gene expression, suggesting that NSD3 is another H3K36me3 reader protein that can direct KDM5B to actively transcribed chromatin (Appendix Figure 3C-D).

KDM5B interacts with many of the same DNA-replication components it targets

Surprisingly, a large set of gene categories from our KDM5B interaction data associate with cell cycle control and DNA synthesis, including DNA replication complex component ORC1 and MCM5/7; kinases involved in cell cycle progression including CDK1 and PLK1; regulators of cohesin and condensin complex such as PDS5A and RCC1/2. At G1/S boundary, the origin recognition complex (ORC1-6) recognizes replication origins and recruits minichromosome maintenance (MCM2-7) complex and other auxiliary proteins for subsequent DNA polymerase loading and replication firing. FACT (facilitates chromatin transcription) is an evolutionarily conserved H2A-H2B chaperone implicated in transcriptional elongation (Belotserkovskaya & Reinberg, 2004). Interestingly, FACT was also shown to stably interact with the MCM complex (Gambus et al, 2006; Tan et al, 2006) and facilitates chromatin DNA replication, presumably through its nucleosomal reorganization activity. Co-IP confirmed that KDM5B physically interacts with components of the ORC, MCM and FACT complex (Figure 3.7C). Importantly, KDM5B *Drosophila* homolog Lid also interacts with the MCM complex (Moshkin et al, 2009), suggesting an evolutionarily conserved binding between KDM5 family members and the DNA replication machinery. We propose that KDM5B can

transcriptionally activate a cell cycle/DNA synthesis-associated gene network and can directly interact with the cell cycle/DNA replication machinery to sustain a highly proliferative state of ESCs permissive to naïve pluripotency (Figure 3.7D).

Discussion

Nanog is required for the initiation and establishment of naïve pluripotency (Chambers et al, 2003; Mitsui et al, 2003; Silva et al, 2009) and is believed to orchestrate epigenetic machinery that can reprogram the somatic cell epigenome (Chambers et al, 2007). In the current study, we identified *Kdm5b* as a Nanog downstream target important for ESC self-renewal and pluripotency via unbiased gain and loss of function studies.

Kdm5b expression is highly enriched in the ICM, pluripotent stem cells and testis (Figure 3.2B-C). Both the ICM and testis are a reservoir for naïve pluripotent cells (Dejosez & Zwaka, 2012). Similar to Nanog, *Kdm5b* null embryos are peri-implantation lethal and *Kdm5b* homozygous mutant ESCs cannot be derived (Catchpole et al, 2011). In addition, Dey and colleagues reported that knockdown of *Kdm5b* in ESC triggered senescence (Dey et al, 2008). We also showed that acute *Kdm5b* knockdown by siRNA compromised ESC proliferation and self-renewal (Xie et al, 2011). In contrast, *Kdm5b* null ESCs could be established via *in vitro* genetic selection (Schmitz et al, 2011). Interestingly, *Nanog* null ESCs can also be generated and maintained *in vitro* (Chambers et al, 2007). It would be interesting to determine whether *Kdm5b* null ESCs show some of the same gene expression and pluripotency deficits as *Nanog* null ESCs.

We show here that *Kdm5b* expression can sustain ESC self-renewal even following long-term culture in the absence of LIF. Importantly, *Kdm5b*-expression cells grown in the absence of LIF maintained pluripotency-marker genes expression and were able to form all somatic lineages in a teratoma assay. Consistent with Dey et al.'s study, these data suggest that KDM5B is important to ESC fate control by stimulating proliferation and repressing differentiation (Dey et al, 2008). Remarkably, we show that

ectopic *Kdm5b* compensates for loss of *Nanog* and is required for *Nanog*-mediated self-renewal, suggesting that *Kdm5b* is a major downstream effector of *Nanog*. This observation echoes the proposal by Niwa and colleagues that *Tbx3*-*Nanog* and *Klf4*-*Sox2* represent parallel pathway downstream of LIF to support ESC pluripotency.

Nanog is required for somatic reprogramming to switch to ground pluripotent state and forced expression of *Nanog* can greatly enhance fusion mediated reprogramming (Silva et al, 2006; Silva et al, 2009). Our data shows that ectopic *Kdm5b* expression can also enhance reprogramming efficiency of NSCs. We are currently testing whether *Kdm5b* can replace *Nanog* to enhance fusion-mediated reprogramming between ESCs and NSCs and whether *Kdm5b* is required for generating ground state iPSCs. In support to *Kdm5b*'s role in somatic reprogramming, the ability of *Nanog* to increase ESCXMEF fusion-mediated reprogramming correlates with a concomitant increase of *Kdm5b* expression (Han et al, 2010). KDM5B is also among the protein fractions from pluripotent cells that can reactivate Oct4-GFP in differentiated cells and boost the reprogramming process (Singhal et al, 2010).

Our previous study demonstrates that KDM5B occupies the transcribed region of highly active genes in ESCs and contributes to productive transcriptional elongation. *Kdm5B* has also been reported to negatively regulate developmental regulator at gene promoters (Schmitz et al, 2011). Thus, *Kdm5b* may have dual roles in regulating pluripotency by functioning as an activator of self-renewal-associated and/or by functioning as gene-promoter bound repressor.

ChIP-Seq data show that KDM5B expression is not enriched at core pluripotency transcription factors and suggest that it does not play a redundant role in the core pluripotency pathway. Interestingly, KDM5B significantly targets a gene-network

involved in cell cycle control and DNA synthesis (Figure 3.6C), suggesting that KDM5B may directly regulate self-renewal. Moreover, epistasis experiments suggest that cell cycle/DNA synthesis genes can partially rescue ESC proliferation and self-renewal following *Kdm5b* knockdown (Figure 3.6D-E). ESCs show atypical cell cycle regulation characterized by an extended period in S phase and a short or undetectable G1 phase (White & Dalton, 2005). A shortened G1 phase has been proposed to facilitate ESC self-renewal by limiting responsiveness to an external differentiation signal (Lange & Calegari, 2010; Singh & Dalton, 2009). During rapid cell cycle progression in ESCs, accurate genetic and epigenetic information is duplicated and inherited. Perturbing the rapid and ordered ESC cell cycle structure should interfere with ESC fate control. Indeed, knockdown of *Cdk2* (Koledova et al, 2009; Neganova et al, 2009), *Cdk1* (Zhang et al, 2010) or overexpressing *p21*, *p15*, *p16* (Ruiz et al, 2010) can slow ESC cell cycle progression, causing ESC to exit self-renewal and irreversibly differentiate. Moreover, modulating cell cycle regulators by ectopic expression of *c-Myc* (Takahashi & Yamanaka, 2006), *Cdk4/cyclinD1* (Ruiz et al, 2010) or repression of Rb (Ruiz et al, 2010), *p53/p21* (Marion et al, 2009), *Ink4/Arf* (Li et al, 2009) can boost somatic reprogramming. These studies suggest that the KDM5B-targeted DNA synthesis gene network described here may play essential roles in ESC self-renewal and pluripotency.

Interestingly, proteomic analysis of KDM5B-associated polypeptides identified proteins directly involved in cell cycle control and DNA synthesis, including pre-replication complex (Pre-RC) components MCMs, ORCs and FACT component SSRP1. Consistent with our data, purification of the *Drosophila* Lid complex also identified an association with multiple MCM proteins (Moshkin et al, 2009), suggesting a conserved association of KDM5 family members with the DNA replication machinery. Moreover, H3K36me3/MRG15 can recruit KDM5B to transcriptionally active chromatin (Xie et al,

2011) and H3K36me3/MRG15 demarcates a distinct chromatin architecture that correlates with higher ORC binding and earlier replication timing (Filion et al, 2010). The physical interaction of KDM5B with DNA replication machinery may function to ensure faithful epigenetic inheritance and to limit exposure to differentiation signals.

Finally, *KDM5B* is also upregulated in multiple human cancers and positively regulates cancer cell proliferation and self-renewal (Hayami et al, 2010; Roesch et al, 2010; Xiang et al, 2007; Yamane et al, 2007). The direct regulation of DNA-replication machinery described here may also contribute to *Kdm5b*'s role in cancer oncogenesis. To gain further insight into how KDM5B regulate gene expression a more detailed understanding of how KDM5B is recruited to chromatin is necessary. Likewise, it is also important to dissect the KDM5B-associated multiprotein complexes and understand their biological function in transcription, splicing, DNA repair and DNA replication in both stem cells and human cancers.

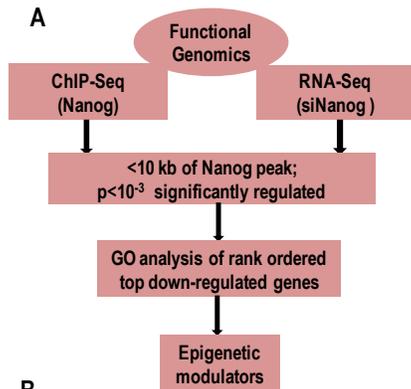
Acknowledgements

We thank Dr. Markus Grompe for providing the Fah immunodeficiency mice and Annelis Haft for her skillful mouse surgery work in the teratoma formation experiment. We thank Dr. Lev Fedorov and Yinming Wang for blastocyst microinjection and embryo dissection. We also thank Dr. Larry David and John Klimek for their help on running and analyzing the KDM5B proteomics samples. Anti-Tyr-P Stat3 antibody (Cell signaling) is a kind gift from Dr. Vince Bicocca from Dr. Brian Druker's lab.

Figures and Figure legends

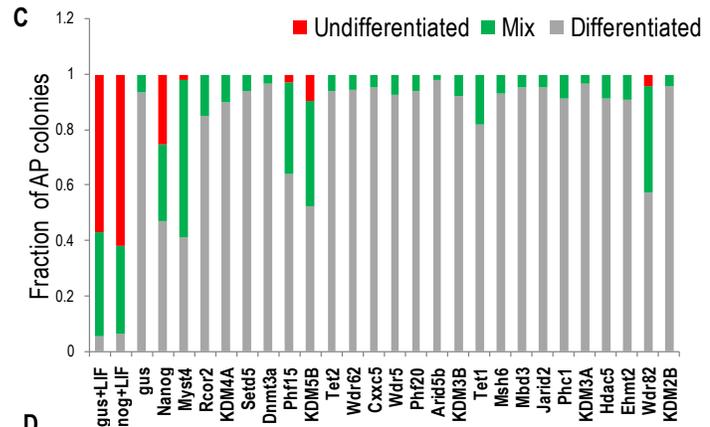
Figure 3.1 Functional genomics assays identify novel Nanog-regulated chromatin modulator genes regulating ESC identity.

A. Flow chart depicting the intersection between Nanog ChIP-Seq and siNanog RNA-Seq libraries. Gene ontology analysis specifically filtered out chromatin modulators that are bound within 10kb of Nanog peaks and significantly down-regulated ($p < 10^{-3}$) by Nanog siRNA knockdown. **B.** Top Nanog-regulated chromatin modulator genes from (A) are rank-ordered by fold change and the top 26 genes are shown using 1.4 fold decrease (-1.4) as cut-off. Red color indicates reported genes that are required for ESC maintenance. Black color indicates genes that have been studied by genetic knockout or siRNA knockdown that do not affect mice early development or ESC self-renewal. Green color indicates novel chromatin modulator genes whose functions have not been studied or not well understood in ESCs. **C.** Mouse ESCs stably expressing indicated Nanog target genes were subjected to LIF withdrawal for 5 days before alkaline phosphatase (AP) positive colonies were quantified. Quantification of undifferentiated, mixed and differentiated colonies based on AP activity staining. **D.** Representative images of AP staining from (C). **E.** Representative AP images of novel chromatin modulator genes that are required for ESC maintenance in RNAi assay.

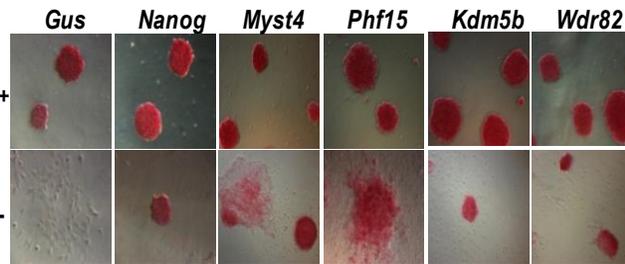


B

	FC	Gene ID	Name and Description
1	-5.57	<i>Myst4</i>	MYST HAT 4
2	-4.68	<i>Rcor2</i>	REST corepressor 2
3	-3.27	<i>Kdm4a</i>	Lysine demethylase 4A
4	-3.07	<i>Setd5</i>	SET domain containing 5
5	-2.99	<i>Dnmt3a</i>	DNA methyltransferase 3A
6	-2.84	<i>Phf15</i>	PHD finger protein 15
7	-2.71	<i>Kdm5b</i>	Lysine demethylase 5B
8	-2.69	<i>Tet2</i>	5-hydroxymethyltransferase
9	-2.66	<i>Wdr62</i>	WD repeat containing 62
10	-2.56	<i>Cxxc5</i>	CXXC finger 5
11	-2.23	<i>Wdr5</i>	WD repeat domain containing 5
12	-2.13	<i>Phf20</i>	PHD finger protein 20
13	-2.11	<i>Arid5b</i>	AT rich interactive domain 5B
14	-2.10	<i>Kdm3b</i>	Lysine demethylase 3B
15	-2.08	<i>Tet1</i>	5-hydroxymethyltransferase
16	-1.96	<i>Msh6</i>	mutS homolog 6
17	-1.90	<i>Mbd3</i>	methyl-CpG binding protein 3
18	-1.89	<i>Jarid2</i>	Jumonji protein 2
19	-1.88	<i>Arid1b</i>	AT rich interactive domain 1B
20	-1.87	<i>Phc1</i>	Polyhomeotic-like 1
21	-1.84	<i>Kdm3a</i>	Lysine demethylase 3A
22	-1.68	<i>Hdac5</i>	Histone deacetylase 5
23	-1.63	<i>Ehmt2</i>	H3-K9 HMT
24	-1.63	<i>Wdr82</i>	WD repeat domain containing 82
25	-1.47	<i>Chd7</i>	Chromodomain DNA binding protein 7
26	-1.40	<i>Kdm2b</i>	Lysine demethylase 2B



D



E

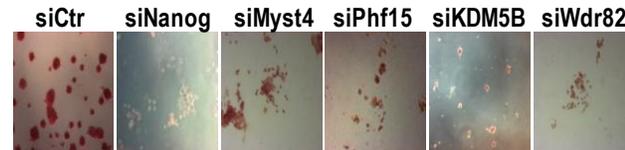


Figure 3.2 *Kdm5b* is highly enriched in pluripotent cells and tissues. **A.** Real time RT-PCR analysis of *Oct4*, *Nanog* and *Kdm5b* expression in pre-implantation embryos at oocytes, 2-cell, morula and E3.5 dpc blastocysts. Data was normalized to *Chuk* gene. Error bar represents Mean \pm SD (n=2). **B.** Immunohistochemical analysis of KDM5B in 3.5dpc pre-implantation mouse embryo shows restricted expression in the inner cell mass. **C.** qRT-PCR analysis of *Kdm5b* cDNA levels in ES cells (J1, CGR8), EC cells (P19), neural stem cells (NSC), terminally differentiated cell lines (MEFs, NIH3T3), indicated tissues and whole embryo day 7, 11, 15 and 17. **D.** Western blot analysis *Oct4*, *Nanog* and KDM5B following RA induced ESCs differentiation at designated time points. Beta-actin is used as a loading control. **E.** Gene expression analysis of KDM5 family members after *Nanog* knockdown.

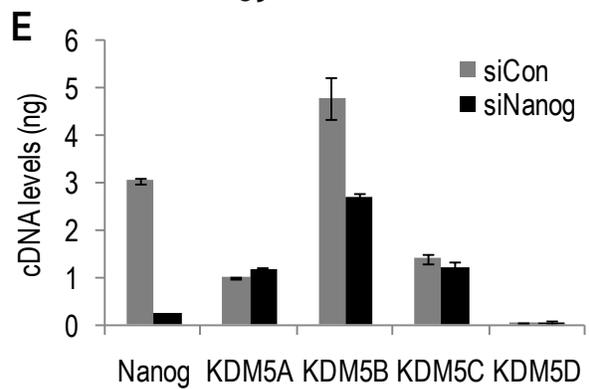
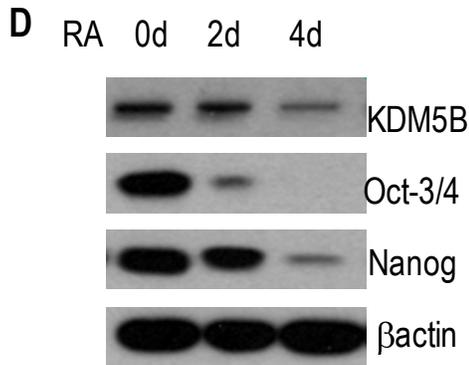
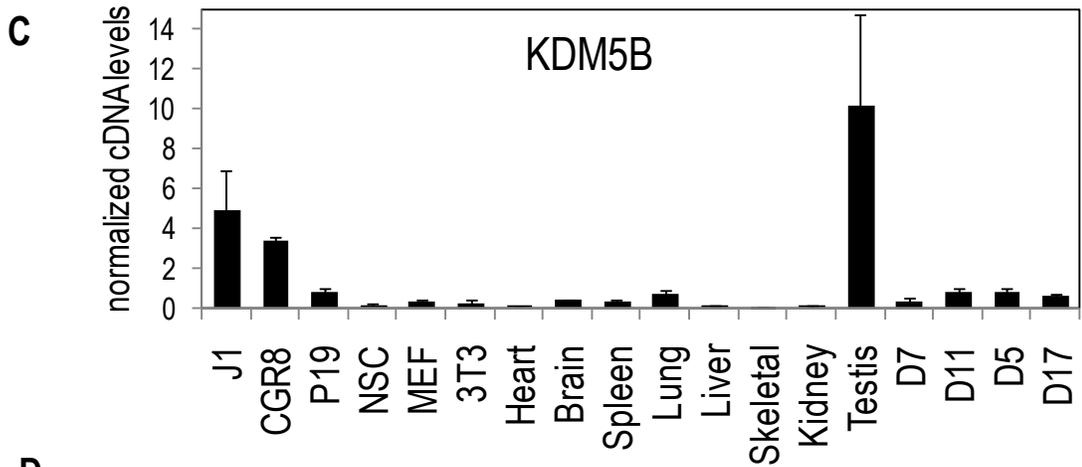
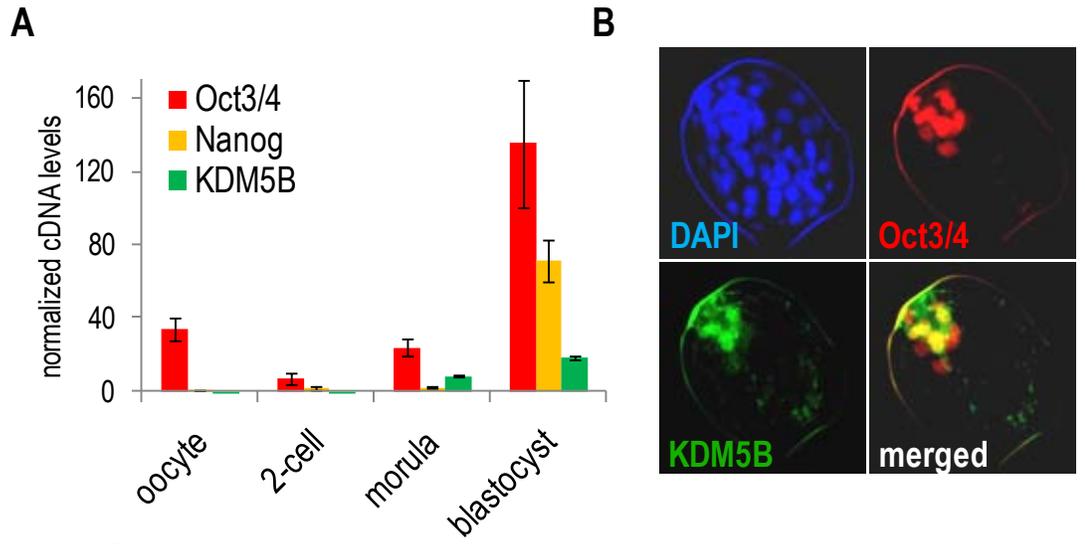


Figure 3.3 *Kdm5b* expression maintains ESC self-renewal in the absence of LIF. **A.** ESCs transfected with *Kdm5b*, *Nanog*, or *Gus* control expression vectors were transiently exposed to medium without LIF for 6 days and quantification of AP positive colonies are shown. **B.** Proliferation assay for individual transgenic ESCs in the absence of LIF. 10,000 ESCs were seeded in 6-well tissue culture plate in triplicates in the presence of puromycin and the number of cells was counted on day 5. Error bars indicate standard deviation (n=3). **C.** qPCR analysis of pluripotency-associated genes in individual transgenic ESCs cultured without LIF for 4 days. The expression level of each transcript in control ESCs cultured with LIF was set to 1. Error bars indicate standard deviation (n=2). **D.** Secondary ESC replating assay for control, *WT-Kdm5b*, *MT-Kdm5b* expressing ESCs after LIF withdrawal for 6 days from (3A). 10,000 cells were replated and cultured in the presence of LIF for 5 days followed by AP staining. Representative images of AP staining from were shown (left) and quantification of AP positive colonies (right).

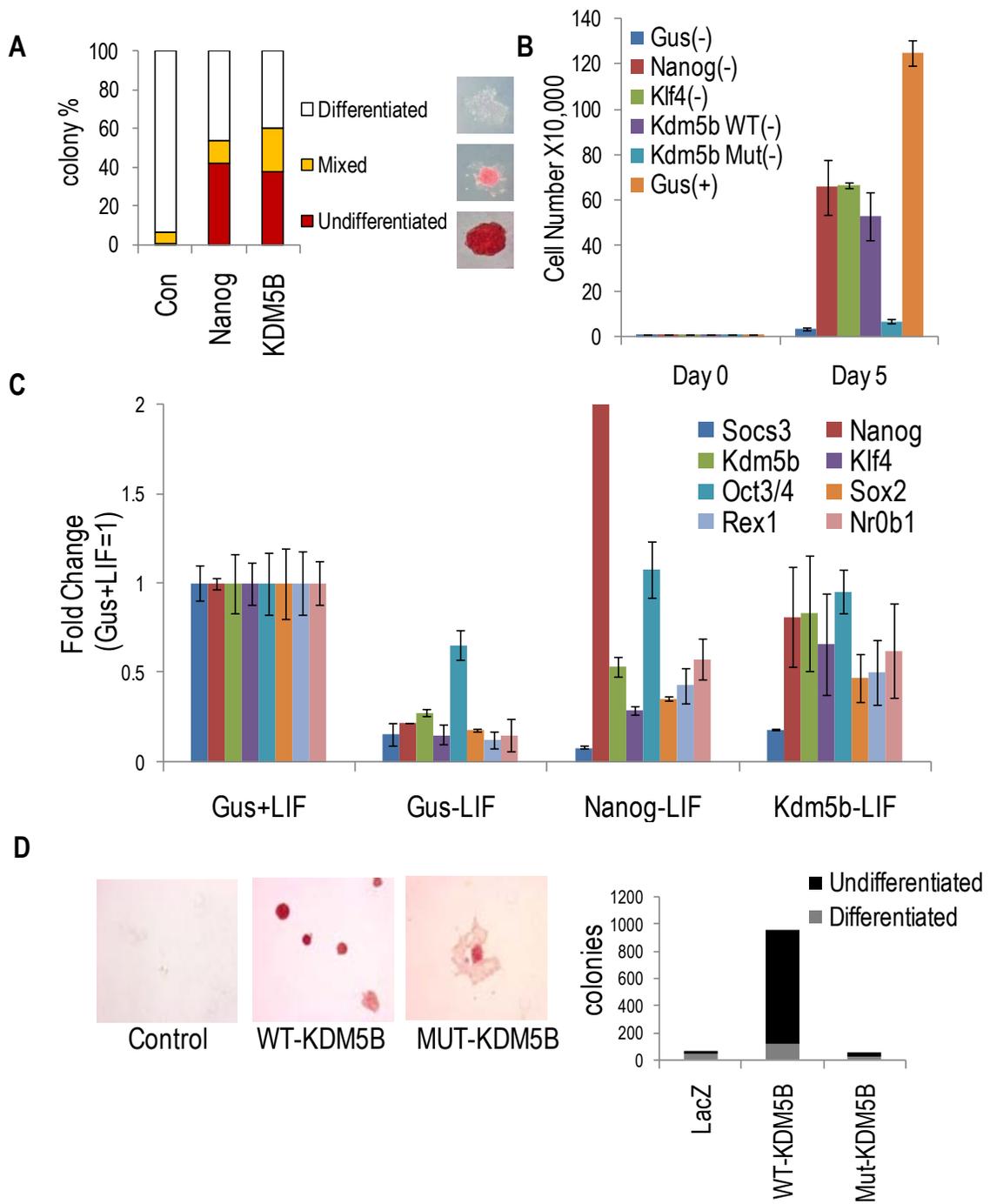


Figure 3.4 *Kdm5b* support ESCs pluripotency in long-term LIF-withdrawal culture

A. ESCs stably expressing Cre recombinase reversible full length *Kdm5b* (E5B) under CAG promoter were continuously passed without LIF for more than 16 passages in 8 weeks. E5B cells were transfected with the pCAG-Cre-IZ vector and selected for zeocin in medium without LIF or with LIF. Representative images of phase contrast morphology and AP staining activity of control ESCs (LIF withdrawal for 6 days), Nanog stable ESCs, E5B cells without LIF at week 8 passage 16, Cre-treated E5BC cells without LIF or with LIF are shown. DsRed fluorescence indicates *Kdm5b* transgene excision. All cells are maintained in medium with 1.5µg/ml puromycin to select for stable transgene expression. **B.** qPCR analysis of pluripotency-associated genes in E5B cells without LIF for 16 passages and 8 weeks, E5B cells treated with Cre without or with LIF for 3 days. The expression level of each transcript in control ESCs cultured with LIF was set to 1. Error bars indicate standard deviation. **C.** Cytogenetic analysis of karyotype test for E5B cells cultured without LIF for 20 passages at week 10 in OHSU Research Cytogenetics Core Laboratory. **D.** E5BC cells at week 10 passage 20 were cultured briefly in LIF medium before injecting into the kidney capsule or subcutaneously in Fah immunodeficiency mice. Same amount of wide type ESCs cultured in LIF were also injected as a control. 3 weeks after injection, tumors were surgically harvested and histological analysis was performed using hematoxylin and eosin staining. Representative images of three germ layers were shown.

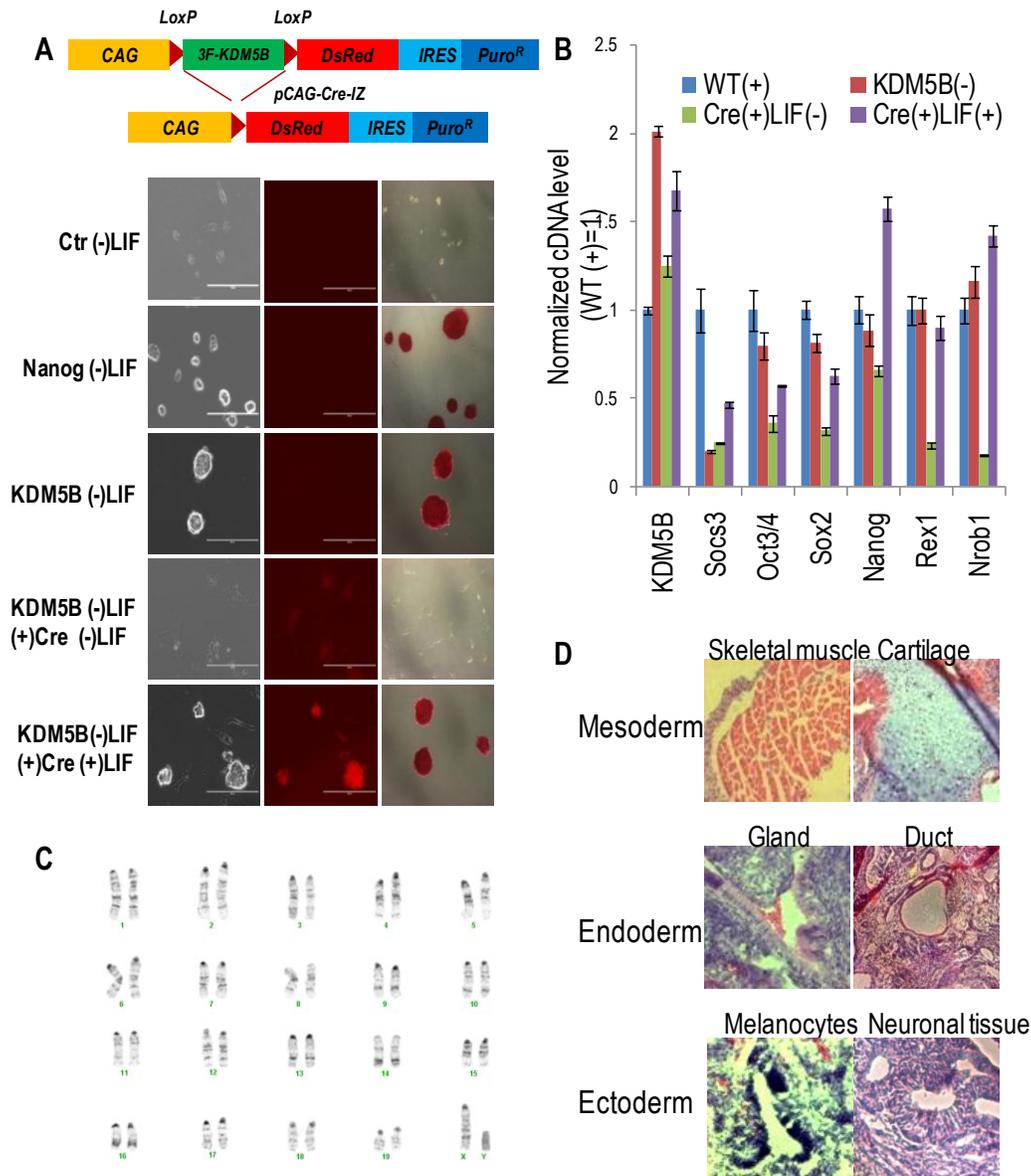


Figure 3.5 *Kdm5b* is downstream effector of Nanog. **A.** ESCs expressing *Kdm5b* or *LacZ* were transfected with control (siCon) or Nanog (siNanog) siRNA. AP positive colonies were imaged 3 days post-transfection. Right panel, percentage of AP positive colonies was quantified. **B.** The expression of lineage-associated markers commonly derepressed by Nanog siRNA treatment was analyzed by quantitative RT-PCR. **C.** ESCs expressing Nanog were transfected with control (siCon) or *Kdm5b* siRNA (siKdm5b). Cells were stained for AP 72 hrs post-transfection. **D.** RT-PCR marker gene analysis of cDNA derived from 5C. **E.** Quantification of iPSC-like colonies based on colony morphology and AP activity from NSCs by retrovirally infecting C-terminal truncated but catalytically active *Kdm5b* (delta C, DC) in combination with *Oct4*, *Oct4+Klf4* (OK). OKM (*Oct4*, *Klf4*, and *c-Myc*) was used as positive control.

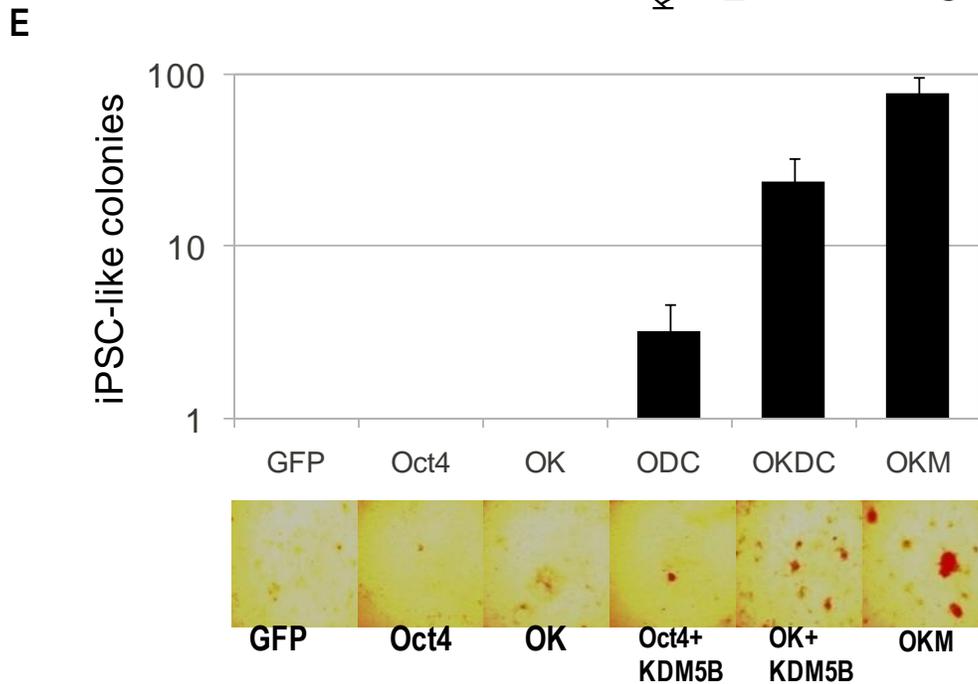
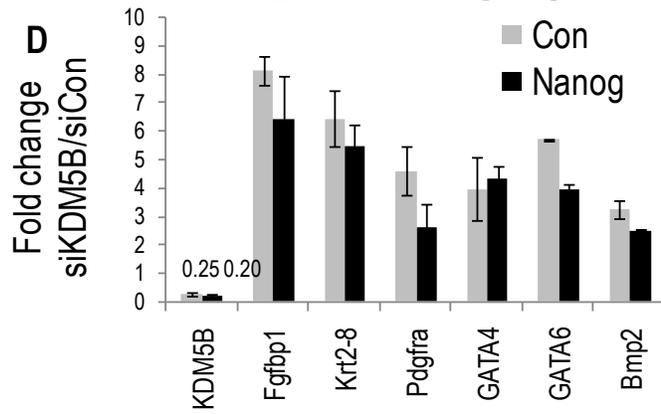
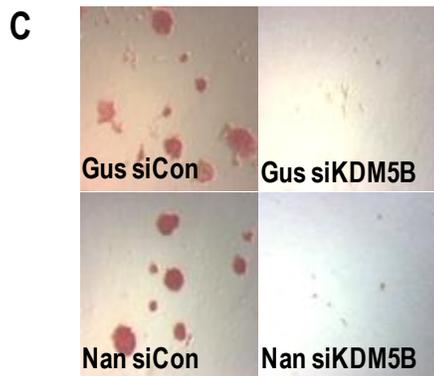
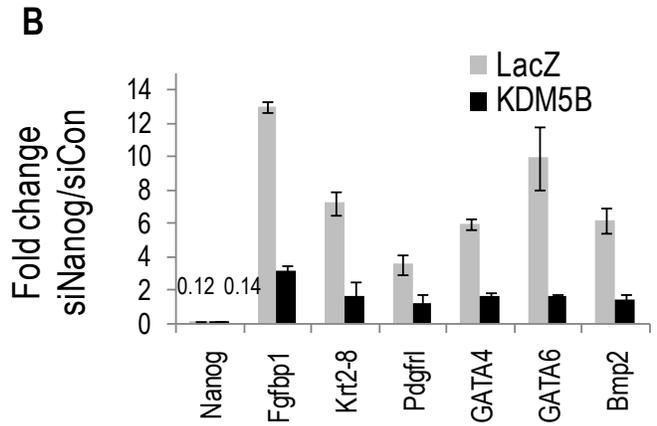
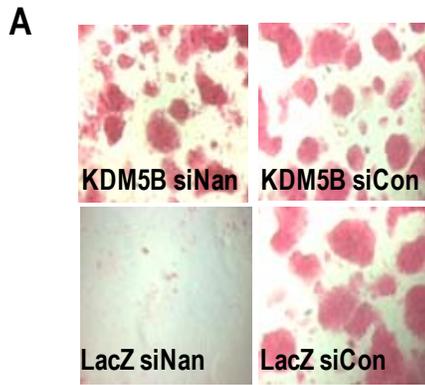


Figure 3.6 Cell cycle and DNA synthesis pathways mediate KDM5B-regulated ESC self-renewal. **A.** LIF was removed from control or *Kdm5b* stable ESC culture for 60 minutes and was added back. Cell lysates at indicated time points were collected and blotted with individual antibodies. **B.** BMP signaling was removed from control or *Kdm5b* stable ESC culture for 24 hours and was added back. Total RNA was extracted at indicated time points and the kinetical expression of BMPs-Smad2/3 target genes *Id1* and *Id3* was analyzed by qRT-PCR. **C.** KDM5B ChIP-Seq and siKdm5b RNA-Seq identified a KDM5B activated gene-network implicated in transcription, chromatin modification, cell cycle progression, DNA synthesis et al. A KDM5B repressed gene network primarily implicated in lineage commitment is also shown. **D.** ESCs expressing *Kdm5b* or *Gus* control were transfected with control (siCon) or *Mcm7* (siMcm7) or *Orc1* (siOrc1) siRNA. AP positive colonies were imaged 3 days post-transfection. **E.** ESCs stably expressing *Gus* control or *Mcm7* or *Orc1* were transfected with control (siCon) or *Kdm5b* siRNA (siKdm5b). 72 hrs post-transfection cells were stained for AP activity and representative images are shown.

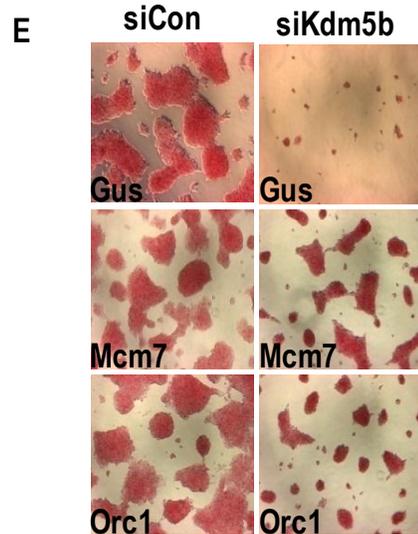
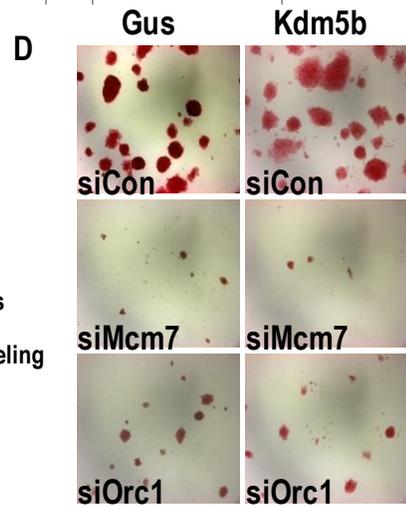
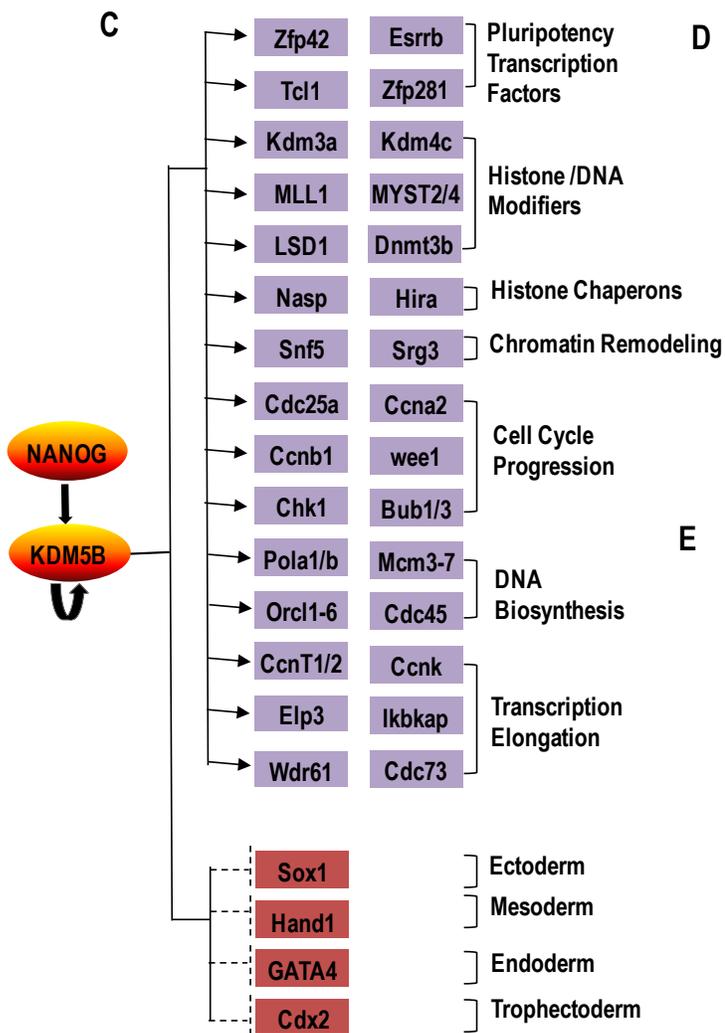
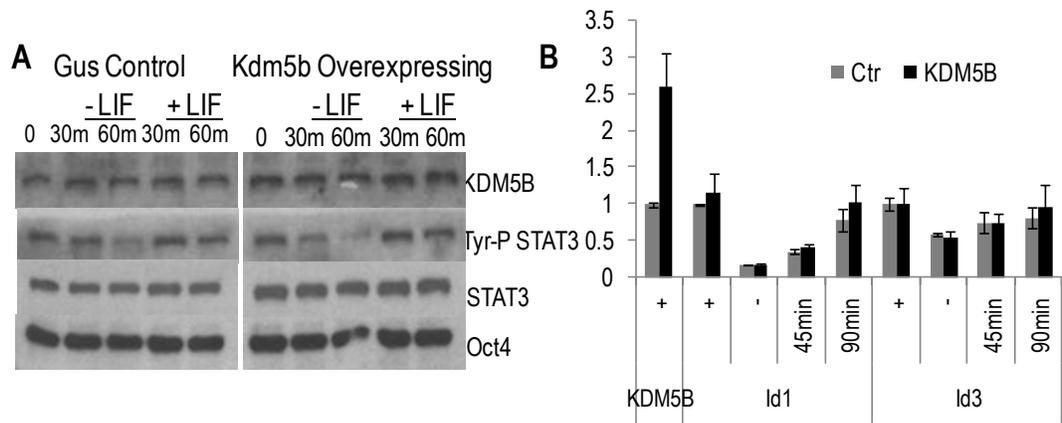
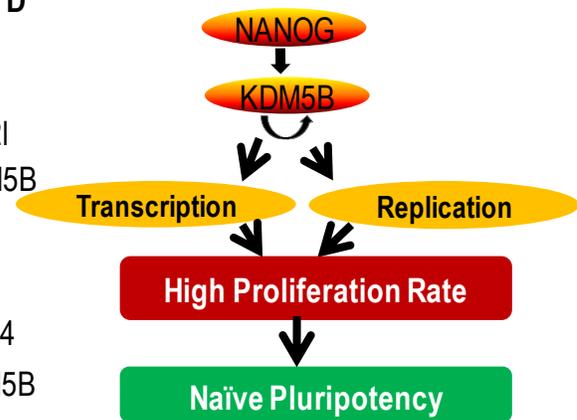
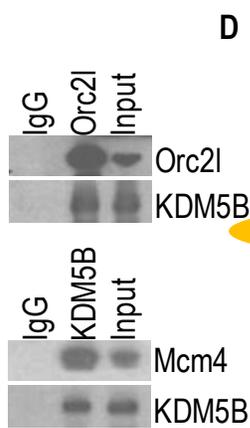
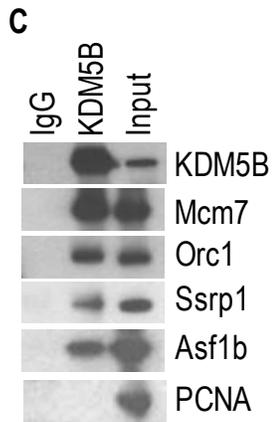
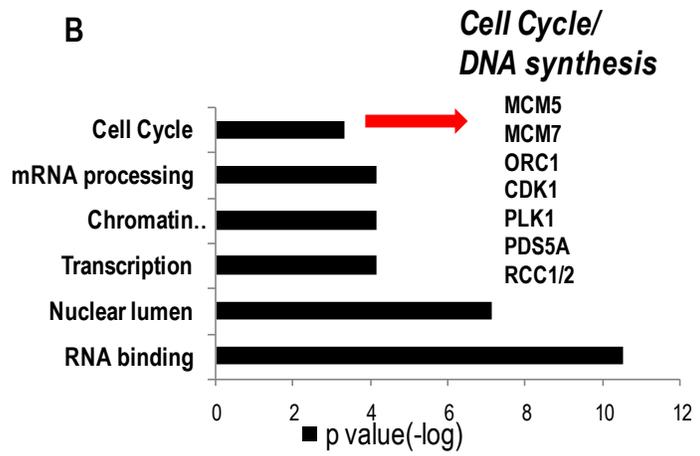
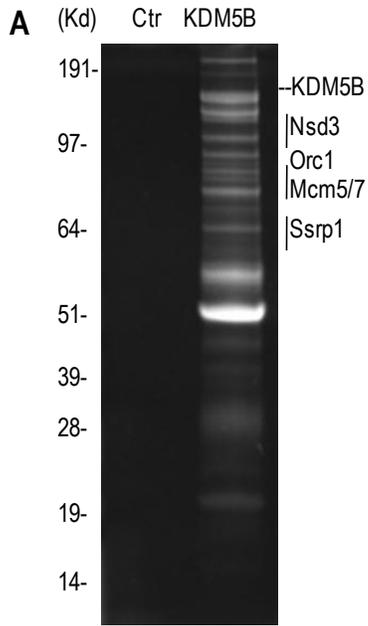
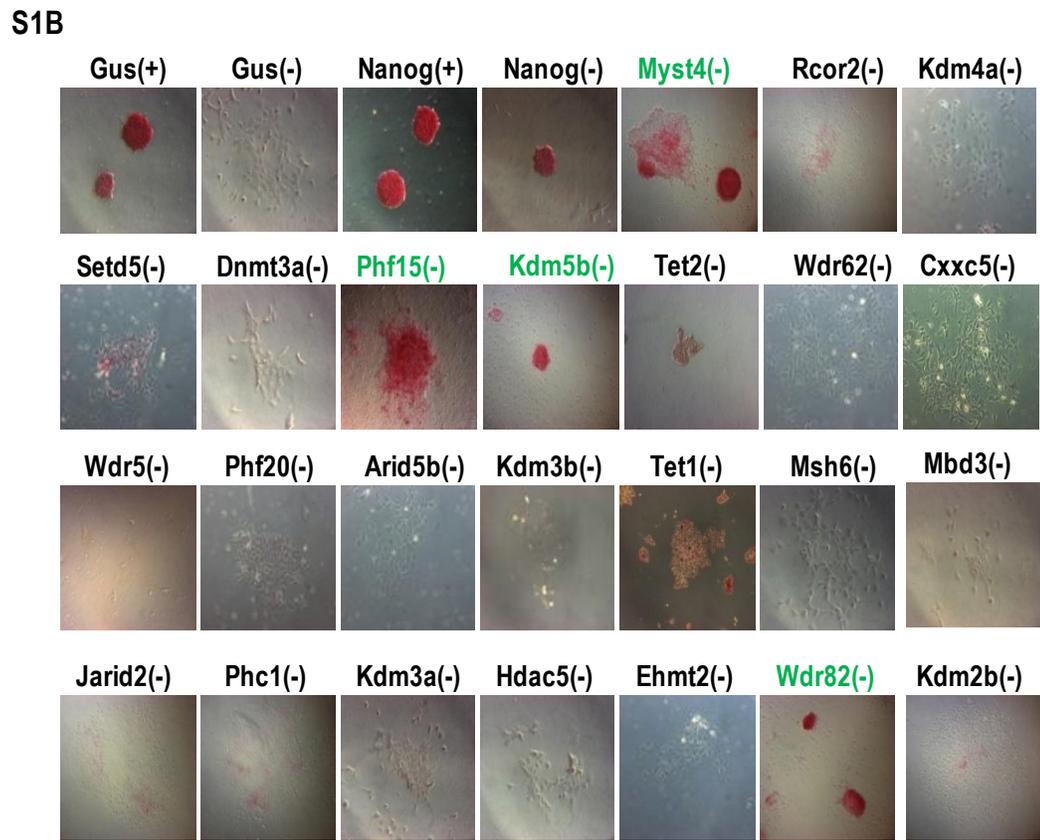
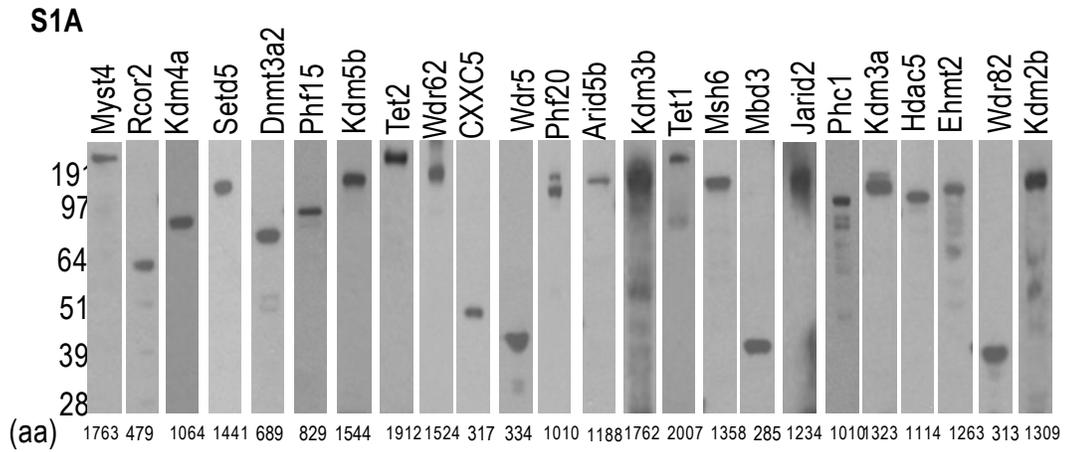


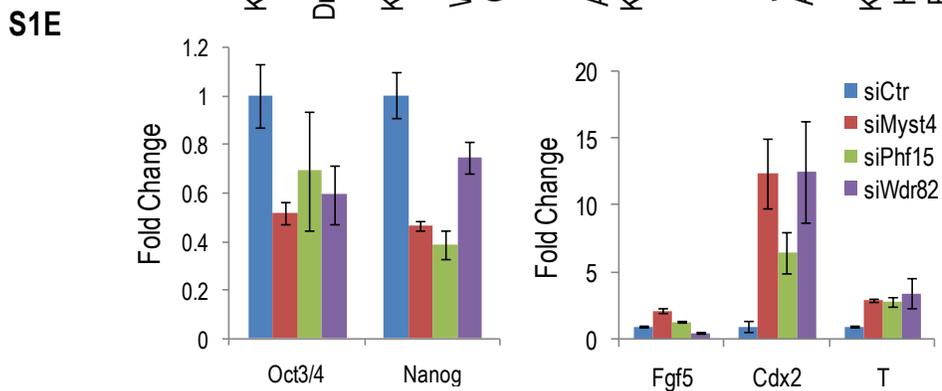
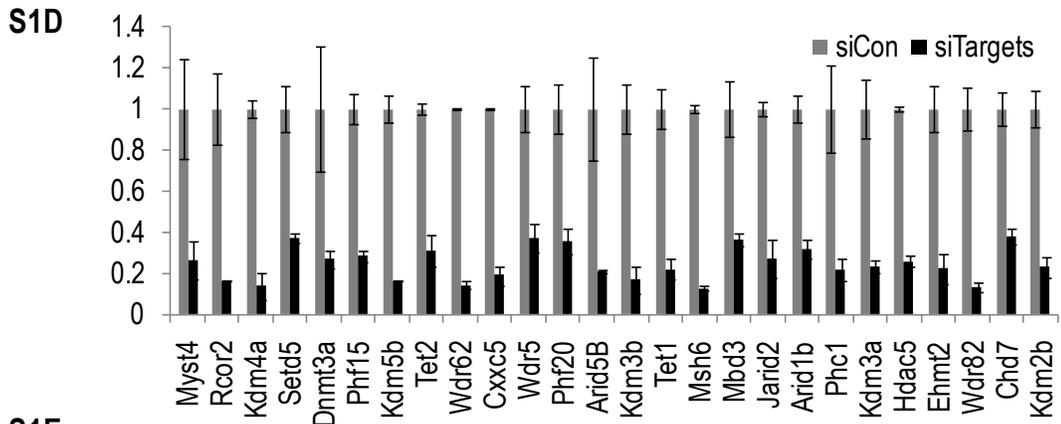
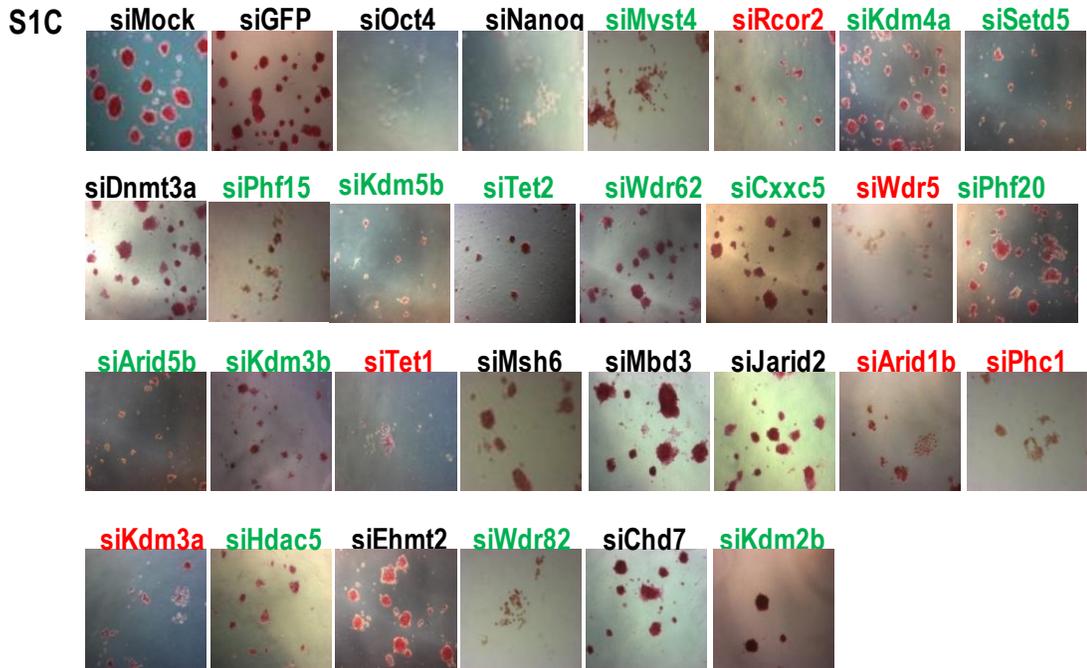
Figure 3.7 Characterizing the KDM5B demethylase complex. **A.** Affinity purification and mass spectrometric analysis of endogenous KDM5B complex. Sypro ruby stain of IgG and KDM5B ESC immunoprecipitates. A complete list of KDM5B-associated polypeptides identified by mass spectrometry is shown in Appendix Figure 2. A large portion of overlapping polypeptides was also identified by Flag antibody-mediated pull down. **B.** Gene ontology analysis of list of peptides with >2 hits from proteomic analyses of KDM5B immunoprecipitates. **C.** Reciprocal Co-IP analysis of KDM5B with proteomics pulled down polypeptides within the cell cycle/ DNA synthesis category. Vectors expressing tagged *Kdm5b* and individual cell cycle/DNA synthesis genes were co-transfected in 293 cells. Immunoprecipitation was carried out using Flag (M2) antibody and western blotting was detected using anti-Myc tag (9E10) antibody. **D.** Proposed model of KDM5B in regulating ESC identify, reprogramming and/or tumorigenesis.



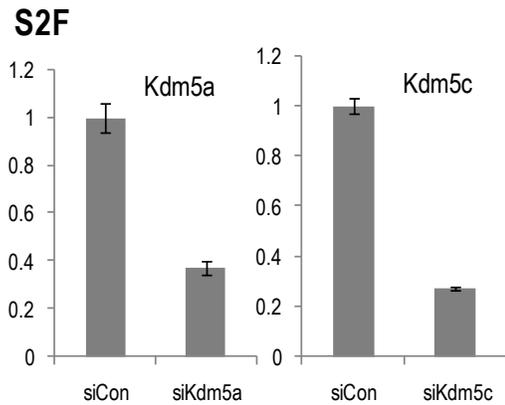
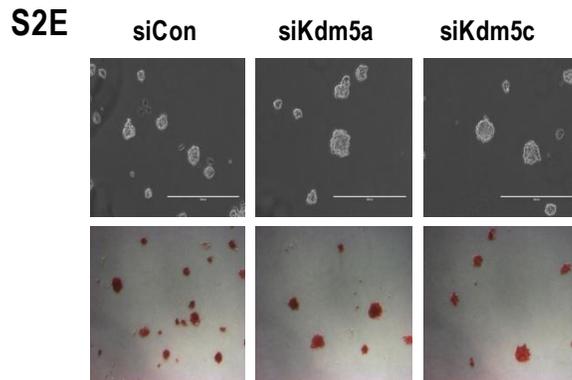
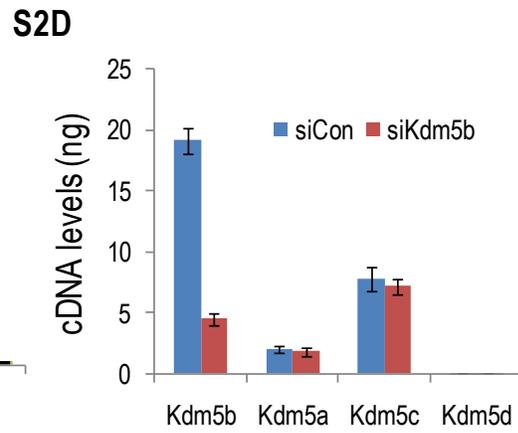
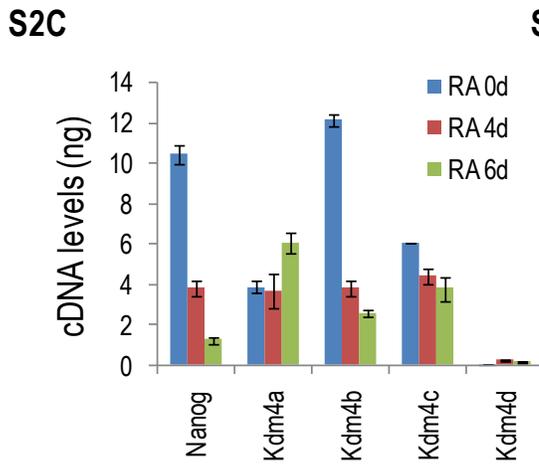
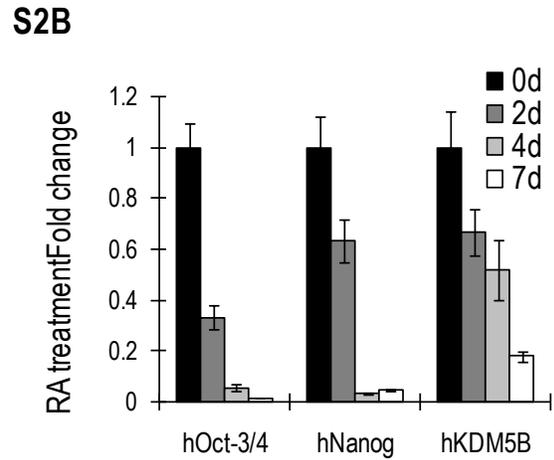
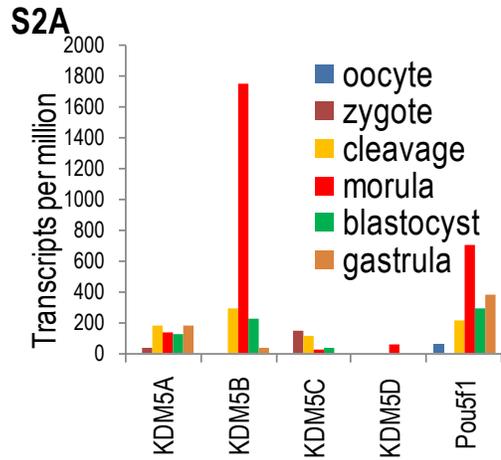
Supplemental Figure and Figure legends

Supplemental Figure S3.1 (S1A) Western blot analysis of stable expression of Nanog-regulated chromatin modulators in E14Tg2a ESCs using anti-Flag (M2) antibody. The predicted protein length encoded by RefSeq genes (amino acid, aa) is indicated at the bottom of each blot. **(S1B)** Individual stable ESC clone corresponding to Nanog-regulated chromatin modifying genes was seeded at clonal density of 50/cm² in duplicates at 6-well tissue culture plate coated with gelatin with or without LIF for 6 days followed by AP staining. Representative images are shown. **(S1C)** Loss-of-function RNAi assay to test the Nanog-regulated chromatin modulator genes in ESC maintenance examined by ESC morphology and AP activity. **(S1D)** RNAi knockdown of *Myst4*, *Phf15* or *Wdr82* decreases pluripotency gene (*Oct3/4*, *Nanog*) expression and increases the expression of lineage --associated marker genes (*Cdx2*, *Fgf5*, *T*).

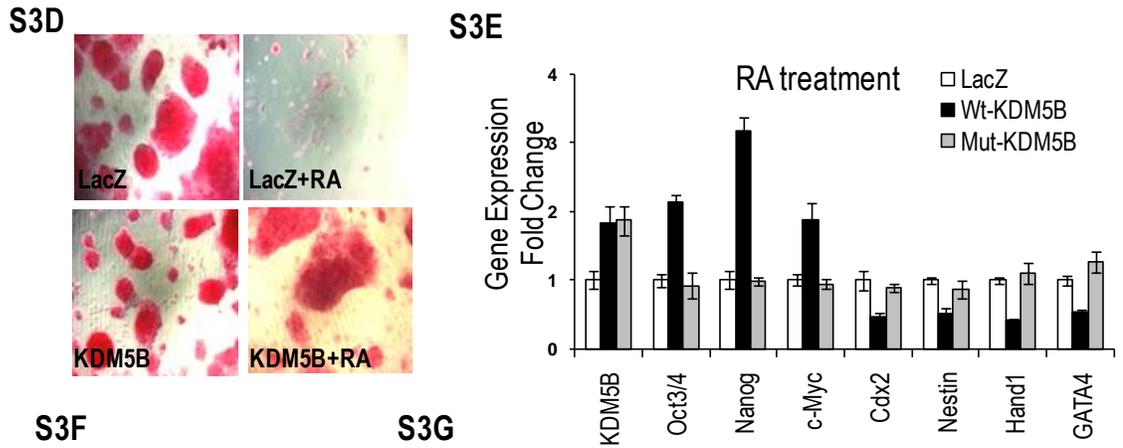
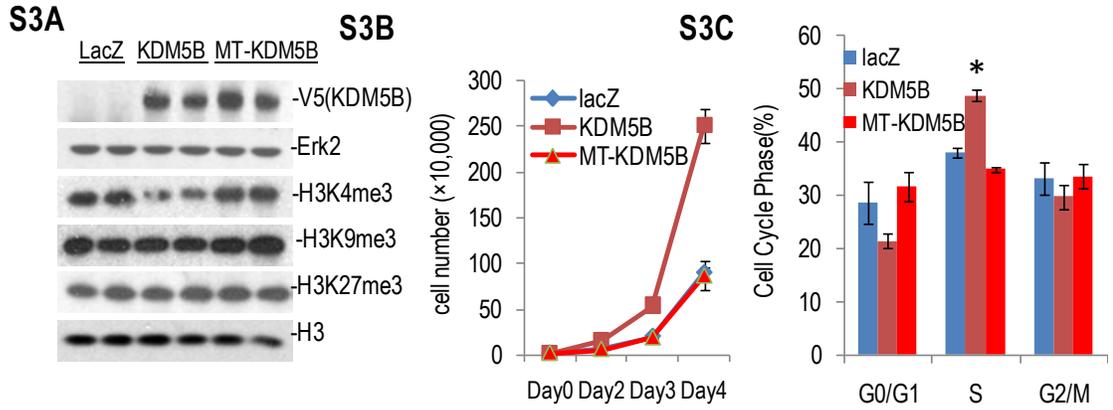




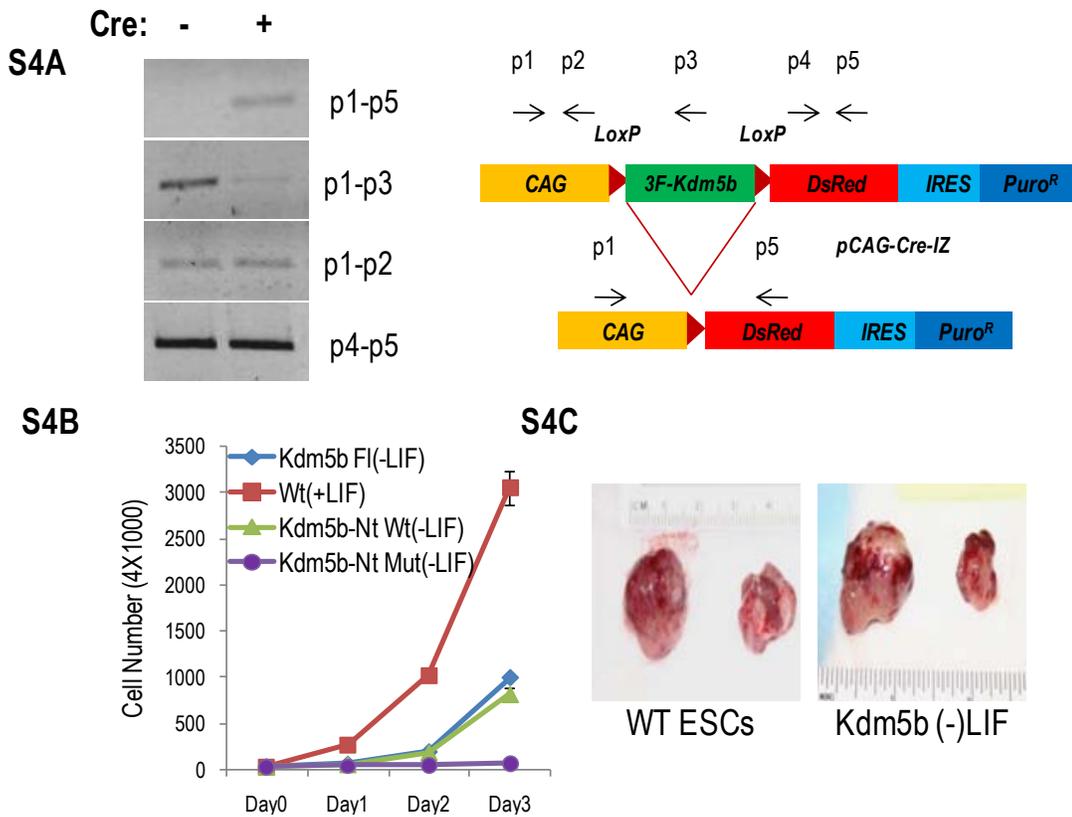
Supplemental Figure S3.2 (S2A) EST profiles of KDM5 family histone demethylases. EST counts (transcripts per million) were collected from NCBI Unigene expression resources (<http://www.ncbi.nlm.nih.gov/sites/entrez>) and plotted for individual KDM5 family member in different developmental stages of mouse early embryogenesis. **(S2B)** qPCR analysis of human *OCT4*, *NANOG*, *KDM5B* at different time points following RA induced human ESCs differentiation. **(S2C)** qPCR analysis of mouse KDM5 family members at different time points following RA induced mouse ESCs differentiation. **(S2D)** qPCR analysis of gene expression of other KDM5 family members in mouse ESCs following acute *Kdm5b* knockdown by siRNA. **(S2E)** Phase contrast and AP staining pattern of mouse ESCs treated with *Kdm5a* or *Kdm5c* siRNA. **(S2F)** *Kdm5a* (left) and *Kdm5c* (right) siRNA efficacy examined by qPCR.



Supplemental Figure S3.3 (S3A) J1 ESCs were transfected with empty vector, wild-type *Kdm5b* (WT-*Kdm5b*), or catalytically inactive *Kdm5b* (MT-*Kdm5b*) and protein extracts were immunoblotted with the indicated antibodies. **(S3B)** J1 ESCs were transfected with control, WT-*Kdm5b*, MT-*Kdm5b* and cell number was counted at the indicated time points. Error bar represents standard deviation (n=3). **(S3C)** FACS analysis of cell cycle phase from J1 ESCs transfected with control, WT-*Kdm5b*, MT-*Kdm5b* expression plasmids. Error bar represents standard deviation (n=3). * denotes $p < 0.05$ using one-tailed student t test. **(S3D)** J1 ESCs transfected with wild type *Kdm5b* or *LacZ* control expression vectors are transiently exposed to medium containing RA for 3 days. Representative images of AP staining are shown (left) with quantification of AP positive colonies (right). **(S3E)** Q-PCR analysis of pluripotency-associated genes and lineage commitment associated genes in control, WT-*Kdm5b*, MT-*Kdm5b* ESCs exposed to RA for 3 days. The expression level of each transcript in control ESCs cultured with RA at day 3 was set to 1. Error bars indicate standard deviation. **(S3F)** Western blot analysis on the protein level of key pluripotency associated transcription factors in control, WT-*Kdm5b*, MT-*Kdm5b* ESCs exposed to RA for 3 days. RNA Pol II and β -actin were used as loading control. **(S3G)** Secondary ESC replating assay for control, WT-*Kdm5b*, MT-*Kdm5b* expressing ESCs treated with RA to induce differentiation. 10,000 cells were replated and cultured in the presence of LIF for 5 days followed by AP staining. Representative images of AP staining from were shown (left) and quantification of AP positive colonies (right). Error bar represents standard deviation (n=2).

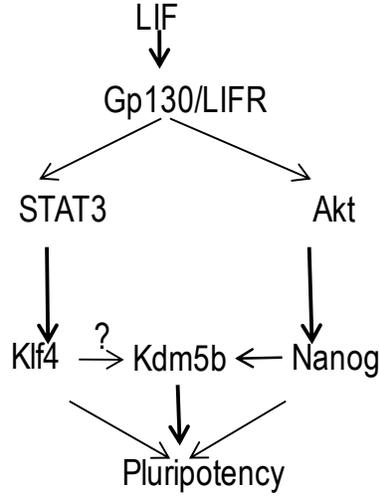


Supplemental Figure S3.4 (S4A). PCR analysis of Cre excision efficiency in E5B cells (passage 16 at week 8). Different combination of primers was illustrated on the right cartoon picture. **(S4B).** Growth rate analysis of E5B cells from passage 16 week 8 compared with ESCs stably integrating wide type and catalytic mutant C-terminal truncated *Kdm5b* (*Kdm5b-Nt*) culture without LIF for 2 weeks. Parental ESCs with LIF are included for comparison. **(S4C).** Gross tumor appearance and size from E5BC cells from Figure 3.4D (passage 20, week 10) compared to that from wide type ESCs with LIF. In each picture, the left side tumor is from subcutaneous injection while the right side from kidney capsule.

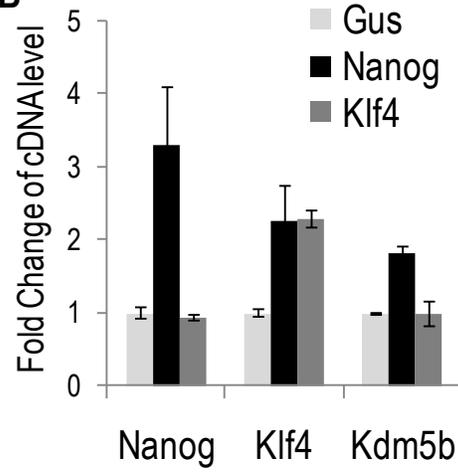


Supplemental Figure S3.5 (S5A) Schematic test of *Kdm5b* in LIF-activated parallel signaling pathway. **(S5B)** Overexpressing *Nanog* but not *Klf4* increases *Kdm5b* expression in ESCs. **(S5C)** ESCs stably expressing *Klf4* were transfected with control (siCon) or *Kdm5b* siRNA (siKdm5b). 72 hrs post-transfection cells were stained for AP. **(S5D)** QPCR marker gene analysis of cDNA derived from S5C. **(S5E)** ESCs stably expressing *Kdm5b* or *Gus* control were transfected with control (siCon) or triple Klf siRNAs (*Klf2+Klf4+Klf5*, siKlfs). AP positive colonies were imaged 3 days post-transfection. **(S5F)** The expression of *Klf4*, *Nanog* and *Klf4*-repressed lineage marker *Fgf5* was analyzed by quantitative RT-PCR following triple Klf siRNA treatment.

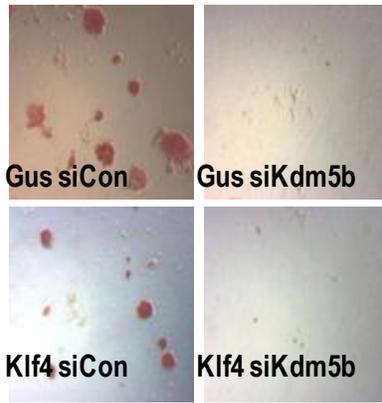
S5A



S5B



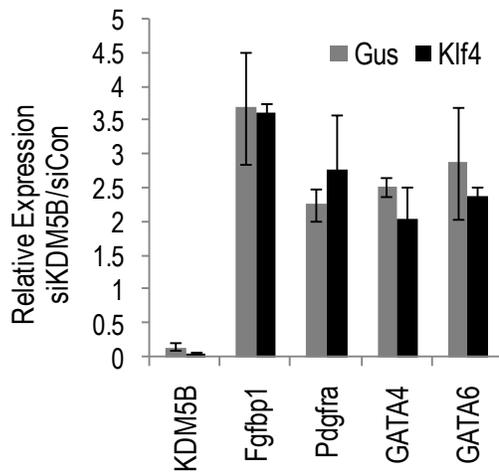
S5C



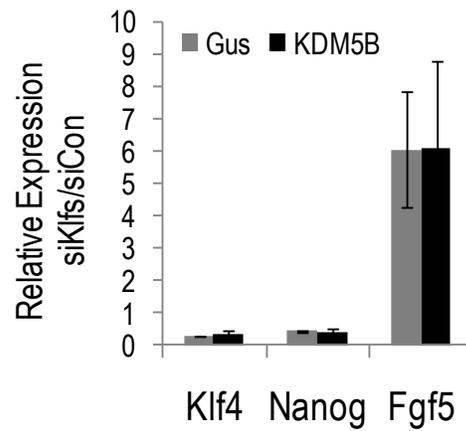
S5E



S5D



S5F



Methods

Cell culture

Feeder-free J1 and E14Tg2a mouse ESCs were cultured at 37°C with 5% CO₂. All cells were maintained on gelatin-coated dishes in Dulbecco's modified Eagle medium (DMEM; GIBCO), supplemented with 10% ESC-tested fetal bovine serum (FBS; New Zealand), 0.1 mM MEM NEAA (Invitrogen, Catalog #11140), 2 mM L-glutamine (Invitrogen, Catalog #25030), 100U/ml/100ug/ml Pen Strep (Invitrogen, Catalog #15140), 0.055 mM -mercaptoethanol (Invitrogen, Catalog#21985), and 1000 U/mL LIF (Chemicon).

Functional Expression Screen

cDNA encoding the top 24 chromatin modifying genes downstream of Nanog were PCR amplified using Phusion DNA polymerase and inserted into pENTR-D/TOPO vector (Invitrogen) to generate Entry clones before subsequently shuttling into pBRCAG-3XFlag-GW-DsRed-IP Destination vector (original CAG expression vector obtained from Dr. Hitoshi Niwa) via Gateway LR reaction (Invitrogen). Individual expression plasmid was transfected into E14Tg2a ESCs to generate stable ES cell lines under the selection of puromycin (1.5ug/ml). Overexpression of individual chromatin modifier genes was examined by western blot by anti-Flag antibody (M2, Sigma). Individual ESC clone for corresponding Nanog-regulated chromatin modifying genes was seeded at clonal density of 50/cm² in 6-well tissue culture plate coated with gelatin without LIF for 6 days before AP staining procedure was performed according to manufacture's instructions (Milipore).

RNA interference screen and transfection

RNAi experiments were performed using shRNA expression vectors as described previously (Xie et al, 2011). Three shRNA sequences were designed for each chromatin modulator genes and pool transfected into ESCs with 1:1:1 ratio using Lipofectamine2000 (Invitrogen).

For transfection, E14Tg2a ESCs were transfected with 2 µg shRNA in 6 well plates or 20 µg in 150mm plates using Lipofectamine2000 or Lipofectamine LTX with PLUS (Invitrogen) according to manufacture's protocol. Puromycin (Sigma) selection was introduced 1 d after transfection at 1.5µg /mL, and maintained for 2-3 days prior to harvesting. Plasmid DNA was also introduced into ESCs using the Nucleofector mouse ES Cell Kit from Amaxa Biosystems (Cologne, Germany Cat. No VPH-1001) according to the manufacturer's instructions and selected with puromycin where indicated.

Teratoma formation assay

E5B cells passaged for more than 8 weeks in the absence of LIF were treated with Cre-recombinase and briefly cultured in ESC medium with LIF. These E5BC cells were mixed with cold matrigel at a density of $1 \times 10^5/\mu\text{l}$ and 2 µl droplet of cells were microinjected into the left kidney capsule or the subcutaneous tissue in immunodeficient C57BL/6 mice (Fah mice as a kind gift from Dr.Markus Grompe's lab). Wide type parental E14Tg2a cells were used as a positive control. Three weeks after microinjection teratomas were harvested and fixed in 4% paraformaldehyde overnight and transferred to 70% ethanol before analysis. Samples were immersed in 30% sucrose in PBS overnight and embedded in O.C.T (Tissue-Tek) under methylbutane-dry ice bath.

Cryosections were obtained using Cryostat and microtomed at a thickness of 10 μm for hematoxylin/eosin staining.

Reprogramming assay

We generated pMXs retroviral constructs that express both full length/C-terminal truncated wild-type and mutant Kdm5b in conjugation with classical *Oct4/Sox2/Klf4/c-Myc* cocktail using standard procedure as described previously (Takahashi & Yamanaka, 2006). NSCs were derived from Oct4-EGFP stable ESCs in the presence of EGF and FGF. All NSCs were negative for EGFP expression. Positive iPSCs were EGFP positive and stained with strong AP activity.

RNA isolation, reverse transcription, and real-time PCR analysis

Total ESCs RNA was extracted using TRIzol (Invitrogen), dissolved in DEPC-treated water (Ambion), and treated with DNase I (Invitrogen) before reverse transcription. First strand cDNA synthesis was carried out with 150ng of total RNA using random priming and M-MLV reverse transcriptase (Invitrogen) at 37 °C for 50 minutes. In all experiments, a no MMLV control was included and no signal was detected. Endogenous mRNA levels were measured by quantitative real-time PCR analysis based on SYBR Green detection with the IQTM5 Multi-Color Real-Time PCR Detection System (Bio-Rad). Quantification of gene expression level was calculated based on the $\Delta\Delta\text{Ct}$ method normalized by housekeeping gene Gapdh or Hprt. Experimental data were presented as Mean \pm SEM between biological replicates. Student t-test was performed to obtain statistical significance.

Mouse embryo culture and Immunofluorescence

Mouse embryos were derived from CD1 mice. Briefly, 3 to 5 week old female mice were superovulated with 5 IU PMSG and 5 IU hCG 46 hours later. Then females were mated with background matched males and checked for the presence of a copulation plug the following morning. Fertilized embryos were collected by harvesting the oviducts into M2 medium (Sigma, Catalog # M7167), tearing the ampulla and releasing them into a hyaluronidase (Sigma, Catalog # H4272)/M2 solution for dissociation from cumulus cells. Embryos were maintained in gassed KSOM (Millipore, Catalog # MR-020P-5F) in a water-jacketed, 5% CO₂ incubator at 37°C and 95% humidity. mRNAs from two groups each with at least 20 embryos at different developmental stages (Oocyte, 2-Cell, Morula, Blastocyst) were isolated by Oligotex® Direct mRNA Mini Kit (QIAGEN, catalog # 72022), followed by reverse transcription via Superscript III (Invitrogen, catalog # 18018-051) and qRT-PCR analysis. For immunochemistry, 3.5d.p.c embryos were fixed in 4% paraformaldehyde for 10 min, permeabilized with 0.1% Triton X-100 for 5 min and blocked with 5% BSA in TBST for 1 hr. Primary antibodies were applied in 5%BSA in TBST overnight at 4 °C as follows: KDM5B 1:500 (Abcam ab50958), Oct3/4 1:200 (Santa Cruz sc-5279). Embryos were then incubated for 1hr with the appropriate Alexa Fluor-conjugated secondary antibodies (Molecular Probes, Eugene, OR) diluted 1:1000 in 5% BSA in TBST. Hoechst staining was used for visualization of nuclei. Fluorescence was visualized using a Zeiss microscope.

Cell cycle analysis

Kdm5b overexpression J1 ESCs were trypsinized, washed with PBS, fixed with 70% ice-cold ethanol and DNA cell cycle analysis was measured by propidium iodide (10 µg/ml PI, Sigma)-stained nuclei using a FACS Calibur machine (Becton Dickinson) in Oregon Stem Cell Center Flow Cytometry Core. Cell cycle compartments were deconvoluted from single-parameter DNA histograms of 20,000 cells and the cell cycle data was analyzed by ModFit program.

Solexa/Illumina next-generation sequencing library preparation

ChIP DNA or cDNA derived from reverse transcription of rRNA-depleted mRNA were prepared for Solexa sequencing as follows: DNA fragments were repaired to blunt ends by the T4 DNA polymerase and were phosphorylated by the T4 polynucleotide kinase using the DNA Terminator Kit (Lucigen). 'A' tailing was performed by adding a single 'A' base to 3' ends with the Exo-minus Klenow DNA polymerase (Epicentre). Double-stranded Solexa genomic adaptors or in-house designed bar code adaptors were ligated to the DNA fragments with T4 DNA ligase (Invitrogen). Ligation products were purified with 1.5 volumes AMPure XP beads (Beckman Coulter) to get rid of unligated adaptors, and subjected to 18 (13-20) PCR cycles of amplification using Phusion Hotstart II DNA polymerase (NEB). Completed libraries were quantified using Qubit 2.0 fluorometer and High Sensitivity DNA quantification kit (Invitrogen). DNA sequencing was carried out using the Illumina/Solexa Genome Analyzer sequencing system (GAII and HiSeq) at OHSU Massively Parallel Sequencing Shared Resource. Sequence reads from each ChIP-Seq or RNA-Seq library are compiled, post-processed and aligned to the reference

genome. ChIP-Seq or RNA-Seq data was verified by real-time PCR. The Impey lab has developed a software pipeline that annotates tags within RefSeq genes and identifies significantly regulated genes using the χ^2 statistic and the Storey Q-test FDR adjustment (Storey & Tibshirani, 2003). Sequence tags that fall outside of RefSeq genes will be analyzed using a sliding-window method with a random background model (Fejes et al, 2008). Custom scripts will be used to identify genes significantly-regulated by *Nanog* or *Kdm5b* knockdown (FDR $p < 0.01$) and these genes will be annotated based on Nanog ChIP-Seq tags or intragenic KDM5B ChIP-Seq tag occupancy. Putative target genes will be selected for further study based on gene ontology and pathway (KEGG2) categories using script-based tools (Gominer, Subpathwayminer) and web-based platforms (David2, ArrayXPath).

Affinity purification of Kdm5b complex followed by mass-spectrometry

For Flag IP, nuclear extracts from 5X14 cm dishes of 3xFlag-*Kdm5b* and control ESCs were prepared (Dignam et al, 1983) and dialyzed to buffer C-100 (20 mM HEPES [pH 7.6], 0.2 mM EDTA, 1.5 mM MgCl₂, 100 mM KCl, 20% glycerol) according to (van den Berg et al, 2010). 60 μ l of anti-FLAG M2 agarose beads (Sigma) were added to 1.5 ml of nuclear extract in No Stick microcentrifuge tubes (Alpha Laboratories) and incubated for 3 hr at 4°C in the presence of 225 units of Benzonase (Novagen). Beads were washed five times for 5 min with buffer C-100 containing 0.02% NP-40 (C-100*) and bound proteins eluted four times for 15 min at 4°C with 0.2 mg/ml FLAG-tripeptide (Sigma). Elutions were pooled and analyzed by mass spectrometry. For immunoprecipitation of endogenous KDM5B complexes, 10 μ g of rabbit polyclonal KDM5B antibody (ab50958, abcam) or IgG were added to 1 ml of nuclear extracts made from J1 ESCs containing

Benzonase overnight at 4°C in No Stick microcentrifuge tubes. Samples were incubated with 100µl Protein A beads for an additional 3 hrs, washed five times for 5 min with C-100* at 4°C, boiled in SDS-PAGE buffer followed by Mass Spectrometry analysis at OHSU Proteomics Shared Resources.

Chapter Four

Summary, Discussion and Conclusion

Summary

Embryonic stem cells (ESCs) derived from the inner cell mass of pre-implantation mouse embryos have an unlimited capacity for self-renewal, can differentiate into all somatic lineages (pluripotency), and are associated with unique chromatin signatures including decondensed euchromatin, hyperdynamic chromatin protein binding, and bivalent histone modifications (Bernstein et al, 2007; Meshorer & Misteli, 2006; Niwa, 2007b). Both embryo-derived and induced pluripotent stem cells show global remodeling of the epigenome including loss of DNA methylation, re-establishment of bivalent histone modifications and re-activation of silenced X chromosomes (Maherali et al, 2007). These studies suggest that a specific epigenetic program is important to establish and maintain the pluripotent state.

The core pluripotency transcription factor *Nanog* is required for naïve pluripotency (Silva et al, 2009; Theunissen & Silva, 2011), a pluripotent state that strictly depends on LIF signaling and can readily contribute to chimaerism (Nichols & Smith, 2009). Moreover, forced *Nanog* expression can support ESCs self-renewal in the absence of LIF (Chambers et al, 2003; Mitsui et al, 2003) and elevated *Nanog* expression promotes transfer of pluripotency after cellular fusion between ESCs and NSCs (Silva et al, 2006). It is believed that *Nanog* choreographs an epigenetic program that erases the somatic epigenome and instills naïve pluripotency. Although *Nanog* was shown to interact with multiple chromatin modifying complexes in ESCs (Liang et al, 2008), the downstream targets of *Nanog* that specify the ESC naïve pluripotency are elusive.

We carried out bioinformatics analyses of *Nanog* ChIP-Seq and si*Nanog* RNA-Seq libraries and identified 26 chromatin modulator genes that are positively regulated

by *Nanog*. Functional overexpression of these chromatin modulator genes was performed to examine their ability to mimic *Nanog* in sustaining ESC pluripotency in the absence of LIF. We also performed a complementary RNAi assay to test the role of the individual chromatin modulator gene in maintaining pluripotency. Both gain- and loss-of-function studies converged on an overlapping set of novel chromatin modulator genes important for ESC identity. We focused on the H3K4me3/me2 demethylase gene *Kdm5b* because *Kdm5b* is one of the top *Nanog* binding targets in two independent *Nanog* ChIP-Seq libraries and because KDM5B is involved in the proliferation and self-renewal of a variety of human cancers (Hayami et al, 2010; Roesch et al, 2010; Xiang et al, 2007; Yamane et al, 2007).

We show that *Kdm5b* expression is highly enriched in the inner cell mass of pre-implantation mouse embryos and pluripotency-associated cells and tissues. Acute depletion of *Kdm5b* compromises ESC proliferation and increases the expression of lineage marker genes. Importantly, forced expression of *Kdm5b*, like *Nanog*, markedly diminishes the requirement of the LIF signaling for ESC maintenance. Moreover, transient overexpression of *Kdm5b* is able to maintain pluripotency-associated gene expression similar to *Nanog* following LIF withdrawal. Stable *Kdm5b* expression can also sustain long-term ESC self-renewal for more than two months in the absence of LIF. These cells are karyotypically normal and can contribute to teratoma formation, suggesting that forced expression of *Kdm5b* confers LIF independent pluripotency. Epistasis experiments demonstrate that overexpression of *Kdm5b* can compensate for loss of *Nanog* and can rescue the loss of pluripotency phenotype. Ectopic *Kdm5b* expression can also boost reprogramming efficiency in conjunction with *Oct4* or *Oct4/Klf4* from neural stem cells. Taken together, we propose that KDM5B can mimic

the action of Nanog to support ESC pluripotency and that KDM5B is a major epigenetic effector downstream of Nanog.

Although KDM5B is believed to function as a promoter-bound repressor, microarray/RNA-Seq analysis of *Kdm5b* knockdown in ESCs shows that it paradoxically functions as an activator of a gene network associated with self-renewal. Unbiased ChIP-Seq reveals that KDM5B predominantly targets to intragenic regions and that it colocalizes with elongating RNA Pol II and H3K36me3, a histone modification marking active transcriptional elongation. We demonstrate that the intragenic occupancy of KDM5B depends on the co-transcriptional deposition of H3K36me3 by H3K36 HMT Setd2 and requires the physical interaction with the chromodomain protein MRG15. Genome-wide MRG15 ChIP-Seq analysis shows high degree of overlap with KDM5B. MRG15 is an ortholog of Eaf3, a component of the yeast Rpd3S complex which functions to repress the cryptic intragenic transcription. KDM5B also physically interacts with the mammalian Rpd3S-like complex components. H3K4me3 ChIP-Seq analysis demonstrates that depletion of *Kdm5b* selectively increases intragenic H3K4me3 at local zones of occupancy. H3K4me3 is believed to recruit the basal transcription machinery component TFIID and to mark active transcriptional initiation (Vermeulen et al, 2007). Consistent with this, knockdown of *Kdm5b* or *Mrg15* significantly increases the cryptic intragenic transcription, promotes the unphosphorylated RNA Pol II recruitment and reduces the elongation-associated RNA Pol II occupancy selectively at KDM5B target genes. We propose a model in which KDM5B activates a unique self-renewal-associated gene network by repressing intragenic cryptic initiation and maintains an epigenetic gradient important for productive transcriptional elongation.

Gene ontology analysis of the intersection of KDM5B ChIP-Seq and si*Kdm5b* microarray/RNA-Seq libraries identified significantly enriched gene categories involved in transcription, cell cycle/DNA synthesis, chromatin modifications and mRNA processing. Interestingly, knockdown of genes involved in DNA replication and mitotic chromosome condensation specifically impaired the maintenance of ESCs but not fibroblast cells (Fazio & Panning, 2010). Therefore, we examined the role of KDM5B-targeted cell cycle control/DNA synthesis genes in ESC identity. Similarly, we found that depletion of the DNA replication machinery components *Mcm7* or *Orc1* impairs the proliferation of ESCs but not fibroblasts in conditions when *Kdm5b* is overexpressed. On the other hand, forced expression of *Mcm7* or *Orc1* can partially rescue the proliferation defect by *Kdm5b* knockdown, suggesting that genes regulating the cell cycle control/ DNA synthesis can partially mediate *Kdm5b* functionality in ESC self-renewal. Surprisingly, proteomic analyses of the KDM5B demethylase complex revealed that KDM5B interacts with components of the DNA replication machinery including components of the the ORC and MCM complex. A similar interaction was reported for the KDM5B homologue Lid in *Drosophila* (Moshkin et al, 2009), suggesting an evolutionarily conserved physical interaction between KDM5 family members and the DNA replication machinery. Therefore, I propose that KDM5B can transcriptionally activate a self-renewal associated gene network and can directly interact with DNA replication machinery to sustain ESCs in a highly proliferative state favorable to naïve pluripotency.

Proteomic analysis of the KDM5B complex identified the H3K36 methylation writer and reader proteins NSD3 and NSD1, elongating RNA Pol II interacting protein THOC1/4, HDAC6, NPM1, and the FACT component SSRP1. In particular, preliminary data shows that NSD3 interacts with both KDM5B and MRG15 and can regulate the recruitment of KDM5B to the intragenic regions of actively transcribed genes. Further

study should reveal the mechanisms by which KDM5B is recruited to intragenic active chromatin and clarify the precise roles of KDM5B during transcriptional elongation.

KDM5B is significantly up-regulated in a wide spectrum of human cancers (Hayami et al, 2010; Xiang et al, 2007; Yamane et al, 2007). One of the hallmarks of tumorigenesis is uncontrolled proliferation (Hanahan & Weinberg, 2011). It would be interesting to determine whether the mechanisms by which KDM5B regulates ESC cell self-renewal also contribute to the self-renewal of human cancers. This hypothesis is being explored via gain and loss of function studies with the eventual goal of developing therapeutic small molecules that target KDM5B.

Discussion

KDM5B is believed to regulate the proliferation and self-renewal of multiple human cancers (Hayami et al, 2010; Roesch et al, 2010; Yamane et al, 2007). In ESCs, forced expression of *Kdm5b* also increases ESC mitotic rate (Dey et al, 2008). We show that *Kdm5b* plays an important role in ESC self-renewal and pluripotency. Moreover, we propose that KDM5B occupies intragenic regions associated with H3K36me3 and activates an ESC self-renewal-associated gene network.

1. Role of *Kdm5b* in ESC proliferation and self-renewal

Kdm5b homozygous knockout embryos are embryonic lethal at the peri-implantation stage and *Kdm5b* null ESCs can not be derived, suggesting that *Kdm5b* plays an important role in early embryogenesis and the specification of pluripotent cells (Catchpole et al). Similarly, the Wynder lab reported that constitutive *Kdm5b* shRNA knockdown ESCs (E14) could not be generated due to cellular senescence (Dey et al, 2008). The Wynder lab also reported that day 9 embryonic bodies (EBs) from gene-trapped *Kdm5b* heterozygous ESCs showed significantly reduced expression of the core pluripotency transcription factors Oct4 and Nanog and derepression of differentiation marker *Egr1* (Dey et al, 2008). These studies are consistent with our observation that acute *Kdm5b* depletion in ESCs compromised their ability to proliferate and self-renew. In contrast, the Helin lab reported that transient *Kdm5b* shRNA knockdown did not alter ESC proliferation or self-renewal (Schmitz et al, 2011). Moreover, Schmitz et al. showed that EBs from *Kdm5b* shRNA knockdown ESCs maintained normal levels of *Oct4* and *Nanog* expression and were defective for the induction of multiple lineage markers (Schmitz et al, 2011). They also reported that transient *Kdm5b* knockdown induced a 2-

fold increase of *Nanog* expression. This result was not seen in two earlier studies (Dey et al, 2008; Xie et al, 2011). The unexpected increase in *Nanog* reported by Schmitz et al. could mask the phenotype observed by other labs. These conflicting observations merit further experimental clarification.

There were several methodological differences between Schmitz et al. and our previous study (Xie et al, 2011). First, Schmitz et al. used E14Tg2a ESCs while we used J1 ESCs. E14Tg2a ESCs were a subclone of E14 ESCs from strain 129/Ola male mouse blastocysts deficient for *Hprt* (hypoxanthine guanine phosphoribosyl transferase) gene. In human males, *HPRT* deficiency causes Lesch-Nyhan syndrome characterized by mental retardation and self-mutilation (Hooper et al, 1987). J1 ESCs were derived from a male agouti 129/terSv strain with normal genotype (Li et al, 1992). Second, Schmitz et al. utilized pLKO viral transductions followed by puromycin selection while we used non-integrating *Kdm5b* siRNA nucleofection for proliferation assays and microarray gene profiling. Rapid siRNA-mediated knockdown also avoids potential homeostatic compensation imposed by antibiotic selection, which could mask phenotypic observation (Freudenberg et al, 2012). Third, in Schmitz et al. ESCs appeared flattened and irregular (Supplementary Figure 1C) while the ESC colonies in our study had a three dimensional dome-like morphology with strong AP activity. It is known that different types of pluripotency exist and that naïve pluripotent ESCs are dome-like round colonies whereas primed pluripotent ESCs (EpiSCs or human ESCs) have a flattened epithelial-like morphology. Mouse EpiSCs express approximately 50% of *Kdm5b* compared to mouse ESCs (Tesar et al, 2007) and proliferated slower than naïve ESCs. In human ESCs, NANOG does not preferentially bind *KDM5B* (Supplementary Figure S2.1A). Schmitz et al did not test whether their cells retain marker gene expression associated with naïve pluripotency. *Kdm5b* siRNA rescue experiments and generation of *Kdm5b*

conditional knockout ESCs from gene trapped *Kdm5b* ESC lines should help resolve the phenotypic differences between Schmitz et al. and our study.

Schmitz et al. also generated conditional *Kdm5b* null ESCs *in vitro* and reported that deletion of *Kdm5b* did not affect ESC morphology, growth rate and pluripotency marker expression (Schmitz et al, 2011). On the other hand, *Kdm5b* knockout mice are embryonic lethal and *Kdm5b* homozygous mutant ESCs could not be derived (Catchpole et al). The basis for this discrepancy is not clear. One possibility is that *Kdm5b* might be required for the establishment of ESC pluripotency but dispensable for its maintenance. Interestingly, *Nanog* null embryos show peri-implantation lethality and transient *Nanog* knockdown in ESCs triggers rapid differentiation (Ivanova et al, 2006). Surprisingly, *Nanog* homozygous mutant ESCs can be generated *in vitro* by genetic deletion (Chambers et al, 2007). This observation led to the hypothesis that *Nanog* is not required for ESC maintenance, but essential for specifying and stabilizing a naïve pluripotent state (Silva et al, 2009; Theunissen & Silva, 2011). The similarity between *Nanog* and *Kdm5b* null phenotypes is potentially consistent with our proposal that KDM5B is a major *Nanog* effector. Therefore, it will be interesting to examine whether *Kdm5b* null ESCs also show germ line maturation defects and gene expression changes as seen in *Nanog* null ESCs. Further studies are needed to clarify whether *Kdm5b* is strictly required for maintenance of self-renewal or establishment and stabilization of naïve pluripotency.

On the other hand, we found that forced expression of *Kdm5b* increased ESC proliferation rate and sustained long-term ESCs self-renewal in pluripotent state in the absence of LIF signaling. This observation is consistent with Dey et al.'s finding that ectopic *Kdm5b* expression can increase markers of mitosis and suppress differentiation-

associated gene expression during EB formation (Dey et al, 2008). Therefore, my work and Dey et al.'s study agree that KDM5B positively modulates ESC proliferation and pluripotency. This is consistent with a wide body of work linking KDM5B to the proliferation and self-renewal in multiple cancer types.

2. KDM5B regulates ESC pluripotency

Kdm5b is one of the top Nanog binding targets identified in two Nanog ChIP-Seq libraries and transcription of *Kdm5b* is positively regulated by Nanog (Xie et al, 2011). Like Nanog, forced *Kdm5b* expression conferred LIF-independent pluripotency. *Kdm5b* overexpression also enhanced the efficiency of somatic reprogramming towards pluripotency. Importantly, we used epistasis experiments to show that ectopic *Kdm5b* expression could support ESC self-renewal even when Nanog is depleted, suggesting that KDM5B is a major epigenetic effector of Nanog. Because Nanog has been suggested to specify and stabilize the naïve pluripotent state, we argue that KDM5B is an important epigenetic regulator of naïve pluripotency. We plan to test whether forced *Kdm5b* expression can, similar to Nanog, promote conversion of primed EpiSCs to naïve ESCs (Silva et al, 2009). It would also be interesting to examine whether ectopic *Kdm5b* expression can push human ESCs into a ground pluripotent state, as has been demonstrated for Nanog-overexpressing hLR5 cells (Buecker et al, 2010). In addition, we plan to examine whether *Kdm5b* depletion would abolish Nanog's ability to promote transfer of pluripotency in fusion-mediated reprogramming (Silva et al, 2006).

3. KDM5B is an activator of gene expression

Nanog is a transcriptional activator that incorporates into an ESC-specific enhanceosome (Chen et al, 2008). To test whether KDM5B is also an activator of

transcription, we carried out ChIP-Seq, RNA-Seq and microarray gene profiling experiments. Our microarray gene profiling and RNA-Seq experiments showed that the majority of significantly regulated genes were down-regulated after *Kdm5b* knockdown in ESCs, suggesting that KDM5B, like Nanog, is an activator of gene expression in ESCs. Although Schmitz et al proposed KDM5B is a repressor of developmental regulators, they surprisingly reported that knock down of *Kdm5b* predominantly decreased, rather than increased the expression of significantly regulated genes in ESCs (Schmitz et al, 2011). Consistent with our proposal that KDM5B is an activator of gene expression, the Swaroop lab identified *Kdm5b* as a downstream target of NRL, the core transcription factor and transcription activator specifying rod photoreceptor cell fate. They reported that the majority of significantly regulated genes were down-regulated after knocking down *Kdm5b in vivo* and were associated with rod photoreceptor function (Hao et al, 2012). It is difficult to reconcile these observations with the Helin lab's proposal that KDM5B functions largely as a repressor.

4. Genome-wide occupancy of KDM5B in ESCs

We reported that KDM5B predominantly occupies intragenic regions of actively transcribed genes associated with ESC self-renewal. Although Schmitz et al. also reported intragenic KDM5B occupancy, they proposed that KDM5B predominantly binds transcription start sites (TSS) and argued that our ChIP-Seq antibody is not specific (Schmitz et al, 2011). However, we have multiple lines of evidence to suggest that the KDM5B antibody (abcam 50958, recognizes N-terminal of KDM5B) we used is specific. First, IP-WB analysis showed a distinct band at the predicted KDM5B protein size (Appendix Figure 4). Second, WB analysis showed that multiple *Kdm5b* siRNA effectively decreased KDM5B protein levels (Appendix Figure 4). Third, *Kdm5b* siRNA

significantly decreased KDM5B occupancy at KDM5B intragenic peak regions (Appendix Figure 4 and Figure 2.2C). Moreover, we showed that KDM5B ChIP-Seq tags predominantly and significantly accumulate at genes down-regulated by *Kdm5b* knockdown (Figure 2.2H-J). The highly significant link between ChIP-Seq occupancy and direct regulation by KDM5B knockdown strongly suggests that our ChIP approach is specific for KDM5B. We have also verified the KDM5B ChIP-Seq signal using an antibody that recognizes a different epitope (Supplemental Figure S2.4B) and on randomly selected intragenic regions (Supplemental Figure S2.4B). To further test the specificity of our KDM5B ChIP-Seq signal at intragenic regions, we plan to obtain the *Kdm5b* inducible knockout ESCs from Schmitz et al. (Schmitz et al, 2011) and examine whether our intragenic KDM5B ChIP-Seq signal is lost following Cre-mediated *Kdm5b* deletion. We have also purchased *Kdm5b* gene trapped ESCs and plan to generate a tetracycline-regulated *Kdm5b* knock out ESC line according to the methodology of Niwa et al. (Niwa et al, 2000). Interestingly, Schmitz et al. also performed KDM5B ChIP-Seq in E14Tg2a ESCs using an in-house generated KDM5B antibody recognizing the C-terminus and reported that 34.8% of KDM5B peaks were found at TSS and 32.7% of peaks intragenically (Schmitz et al, 2011). They reported that H3K4me3 was selectively increased at KDM5B target genes both at TSS and intragenic regions. These results are consistent with our finding that KDM5B has intragenic occupancy. We recently sequenced our KDM5B ChIP-Seq library more deeply and found additional evidence of KDM5B occupancy at promoter regions (Appendix Figure 5), consistent with our proposal that KDM5B may occupy both promoter and intragenic regions. Additional KDM5B ChIP-Seq using different antibodies that recognize distinct KDM5B epitopes revealed that KDM5B occupies both intragenic and TSS regions (Appendix Figure 5). Moreover, the dual localization of KDM5B (promoter and gene bodies) is anticipated

based on its evolutionarily conserved interaction with MRG15 (Hayakawa et al, 2007; Lee et al, 2009b; Xie et al, 2011). MRG15 incorporates into both the promoter-associated NuA4 complex (Doyon et al, 2004) and the gene-body associated Rpd3S complex (Lee et al, 2009b). It is likely that KDM5B may have dual roles in regulating pluripotency: gene-body bound activator of self-renewal associated genes by facilitating transcriptional elongation as well as gene-promoter bound repressor of differentiation associated genes by inhibiting transcriptional initiation. The difference in KDM5B enrichment at promoter and gene bodies in the two studies may reflect the presence of KDM5B in distinct protein complexes *in vivo*. Accordingly, the usage of different epitope-specific antibodies may reveal non-overlapping genomic occupancy (Schmitz et al, 2011). We plan to test this hypothesis by performing KDM5B ChIP-Seq using anti-V5 antibody directed towards both N terminal and C terminal tagged KDM5B in ESCs.

Multiple lines of evidence suggest an intragenic localization of KDM5B. First, functional genomic analysis showed that accumulation of intragenic KDM5B peak tag density positively correlated with rank-ordered down-regulated KDM5B target genes following *Kdm5b* depletion using microarray analysis (Figure 2.2J). We recently confirmed this association using RNA-Seq with different *Kdm5b* siRNAs. Non-specific intragenic occupancy would not produce this genome-wide correlation. Second, both Lid (Lee et al, 2009b) and KDM5A (Hayakawa et al, 2007) are integral components of the MRG15-Rpd3S like complex in *Drosophila* and mammalian cells, respectively, suggesting an evolutionarily conserved partnership of KDM5 family members with MRG15. Reciprocal Co-IP showed that KDM5B bind to MRG15 in both ESCs and heterologous 293 cells (Figure 2.5A-B). Third, MRG15 is an evolutionarily conserved chromodomain-containing protein that specifically recognizes H3K36me3/me2 (Carrozza et al, 2005; Joshi & Struhl, 2005; Keogh et al, 2005; Zhang et al, 2006) and genome-

wide mapping of MRG15 binding revealed that MRG15 significantly occupied coding regions of active chromatin and correlate with H3K36me3 (Filion et al, 2010; Sural et al, 2008; Xie et al, 2011). Moreover, mutation of the MRG15 Drosophila homologue MSL3 chromodomain abolished its ability to recognize H3K36me3 (Sural et al, 2008). Fourth, MRG15 ChIP-Seq analysis revealed highly similar genome-wide binding pattern as KDM5B (Figure 2.5I-K and Appendix Figure 5). Depletion of MRG15 in ESCs significantly decreased intragenic KDM5B occupancy, selectively increased H3K4me3 at KDM5B binding peaks, and decreased the expression of KDM5B activated target genes (Figure 2.5E-G). Fifth, both DRB-mediated inhibition of transcriptional elongation or knockdown of H3K36me3 HMT Setd2 significantly and selectively decreased KDM5B occupancy at intragenic regions of activated target genes (Supplementary Figure S2.7C and Figure 2.4G-H). Setd2 siRNA knockdown also increased intragenic H3K4me3 at KDM5B target peak regions (Supplementary Figure S2.7A). Taken together, these studies support our proposal that KDM5B is localized to intragenic regions associated with active transcription.

5. Biological significance of KDM5B intragenic occupancy

We showed that KDM5B can incorporate into a mammalian Rpd3S-like complex and occupy intragenic regions of actively transcribed genes associated with ESC self-renewal. Whether KDM5B yeast homologues interact with the canonical Eaf3-Rpd3S complex is not known. Genetic deletion of Eaf3, Set2, Rco1, Sin3 all resulted a robust increase of H3K4me3 at the *FLO8* ORF region (Carrozza et al, 2005), suggesting that KDM5 family-mediated intragenic H3K4me3 demethylation may be evolutionarily conserved. In mammalian cells, knockdown of *Setd2*, *Mrg15* or *Kdm5b* all led to a robust increase in H3K4me3 at zones of KDM5B occupancy (Figure 2.3D-E, 2.5F,

Supplementary Figure S2.7A). The increase of intragenic H3K4me3 selectively at KDM5B intragenic peaks further supports KDM5B gene-body occupancy. Functional genomic studies documented that H3K4me3 marks active gene promoters and H3K4me3 could selectively anchor the basal transcriptional machinery component TFIID for transcriptional initiation (Barski et al, 2007; Vermeulen et al, 2007). Accordingly, we proposed that KDM5B could suppress intragenic H3K4me3/me2, cryptic transcriptional initiation, and could play a role in sustaining productive transcription elongation.

Transcriptional elongation co-transcriptionally engages other nuclear processes including pre-mRNA splicing, poly-adenylation and mRNA export. Intragenic KDM5B occupancy suggests that KDM5B may regulate alternative splicing, antisense transcription, or non-coding RNAs. Luco and colleagues demonstrate that SETD2-dependent H3K36me3 marks exon IIIb of FGFR2 in mesenchymal cells, recruits MRG15 and polypyrimidine tract-binding protein PTB to suppress exon IIIb inclusion (Luco et al, 2010). Interestingly, H3K4me3 remains low at the IIIb exon in mesenchymal cells (Luco et al). It remains to be determined whether KDM5B or other KDM5 family members could regulate alternative splicing via their evolutionarily conserved interaction with MRG15. Antisense transcription can interfere with splice site choice by forming double stranded RNAs with sense transcripts (Morrissy et al, 2011). We also showed that some of the intragenic cryptic transcripts induced by *Kdm5b* or *Mrg15* knockdown originated from antisense strand (Supplementary Figure S2.10). Moreover, we found that KDM5B occupied the 3' region of many developmental regulator genes including *Cdx2* and *Sox1*. Preliminary data showed that knockdown of *Kdm5b* increased the expression of the sense transcripts of *Cdx2* and *Sox1* with a concomitant decrease of corresponding antisense transcripts. Our observation echoes the study in fission yeast that shows the Rpd3S complex mutation derepresses antisense production (Nicolas et al, 2007).

Therefore, antisense regulation may be an evolutionarily conserved function of the Rpd3S complex. Recently, Dr. Eric Lander's group reported that KDM5B interacts with several long non-coding RNAs (Guttman et al, 2011). It should be noted that many RNA-binding proteins have also been linked to transcriptional regulation.

The ability of KDM5B to activate a self-renewal associated gene network via suppression of intragenic cryptic transcription may provide a partial explanation for our observation that ectopic *Kdm5b* expression can enhance somatic reprogramming. *Kdm5b* overexpression in somatic cells may promote transcriptional elongation of self-renewal associated genes and generate a chromatin landscape favorable to the pluripotent state. We do not rule out the possibility that KDM5B binding at promoters and suppressing developmental regulators may also contribute to acquisition of pluripotency. The detailed molecular mechanisms await future studies.

6. Specificity of KDM5B-regulated gene network

Gene ontology analysis of KDM5B ChIP-Seq and siKdm5b microarray/RNA-Seq identified highly similar gene categories enriched in broad metabolic processes including 'nuclear lumen', 'nucleotide binding', 'DNA metabolic', 'cell cycle', 'RNA splicing' and 'DNA repair' (Figure 2.2A and 2.2D). Importantly, genome-wide KDM5B occupancy gradient correlates with rank-ordered Ref-Seq genes that are downregulated after *Kdm5b* knockdown (Figure 2.2J). This genomic association suggests that KDM5B occupancy density anticipates the transcriptional outcome of KDM5B target genes, a pattern similar to the elongating RNA Pol II and H3K36me3 (Figure 2.2K-M, Figure 2.4C-E). We propose that KDM5B regulates a specific set of actively transcribed genes in ESCs, many of which are important for ESC maintenance and identity.

Although we propose that H3K36me3 contributes to the docking of KDM5B to actively transcribed chromatin, we did not find perfect overlap suggesting that other mechanisms for KDM5B recruitment exist. A recent study proposed that chromatin proteins can define five principle chromatin types in *Drosophila* (Filion et al, 2010). Active euchromatin can be classified into two subtypes—BRM-marked RED chromatin and MRG15-marked YELLOW chromatin. Unexpectedly, MRG15 enriched chromatin is specifically enriched for H3K36me3, while H3K36me3 is largely absent on BRM-marked RED chromatin, even though transcriptional activity is similar in the two chromatin types. Strikingly, the two active forms of chromatin mark different types of genes. Genes on MRG15-enriched chromatin are associated with metabolic pathways (DNA metabolic processes, structural molecule activity, DNA repair etc.) while genes on BRM-enriched chromatin are linked to signaling pathways (receptor binding, signal transduction, defense response etc.). Given that the *Drosophila* KDM5 homologue Lid forms a stable complex with MRG15 (Lee et al, 2009b; Moshkin et al, 2009), it is conceivable that Lid is also recruited to H3K36me3/MRG15 enriched chromatin. In support to this notion, gene ontology analysis from KDM5B targeted and regulated genes in ESCs shows very similar gene categories as MRG15-demarcated chromatin in *Drosophila*. These results support our model that KDM5B demarcated chromatin regulates specific set of genes important for ESC maintenance.

The differential enrichment of H3K36me3 on MRG15-marked and BRM-marked euchromatin is surprising given that both chromatin types have similar transcriptional activity (Filion et al, 2010). H3K36me3 is believed to be co-transcriptionally deposited by elongating RNA Pol II (S2P) that physically interacts with H3K36 HMT Setd2 (Wagner & Carpenter, 2012). It has been reported that some genes do not require Ser2P for productive elongation. We speculate that three intrinsic gene features may explain the

differential H3K36me3 marking on the two types of euchromatin and thus the gene specific regulation of the H3K36me3-MRG15 pathway.

In yeast, the Set2/Rpd3S pathway appears to target to infrequently transcribed long genes to repress cryptic transcriptional initiation for accurate transcription (Li et al, 2007b). The passage of RNA Pol II along long genes involves multiple rounds of configuration of the transcribed nucleosomes through histone eviction and redeposition (Workman, 2006). The principle function of the Set2/Rpd3S pathway in yeast is to restore the chromatin status (histone acetylation) behind transcribing RNA Pol II (Li et al, 2007b). Thus, long genes have higher probability on the requirement for the S2P RNA Pol II associated-Set2/Rpd3S pathway to safeguard the transcriptional fidelity. It would be interesting to examine whether MRG15-enriched YELLOW chromatin in *Drosophila* is enriched for long genes compared to BRM-marked RED chromatin. It is also known that many genes do not require S2P for productive elongation but instead rely on S5P (Hargreaves et al, 2009; Saha et al, 2011). Interestingly, many of these genes play important signaling roles and are represented in the RED chromatin. It has been reported that H3K36me3 is significantly enriched in intron-containing genes compared to intronless genes, regardless of transcriptional activity (de Almeida et al, 2011). Inhibition of splicing can decrease Setd2 occupancy and H3K36me3 (de Almeida et al, 2011). Because genome-wide H3K36me3 does not completely mirror that of S2P RNA Pol II, de Almeida et al. proposed that co-transcriptional spliceosome assembly and splicing additionally enhance recruitment of Setd2, leading to enrichment of H3K36me3 associated with splicing (de Almeida et al, 2011). Taken together, gene length, gene function and splicing requirement could all contribute to the deposition of H3K36me3 to specific subsets of genes. It is conceivable that intron-containing long genes carrying specific set of functions have evolved to require the coating of H3K36me3 histone mark

to maintain the fidelity of transcription elongation and splicing. One prediction of this hypothesis is that KDM5B regulated genes are characterized by these gene features, which will be examined in further studies.

7. Potential role of KDM5B in DNA replication

Gene ontology analysis of KDM5B regulated genes from both the *Kdm5b* knockdown microarray and RNA-Seq experiments revealed highly overlapping sets of genes involved in cell cycle control and DNA synthesis. We and other labs demonstrated that knock down of many of these KDM5B target genes produced an ESC-specific proliferation defect (Fazzio & Panning, 2010), suggesting they are important to ESC self-renewal. Ectopic expression of genes involved in DNA replication could partially rescue the proliferation defect after *Kdm5b* knock down (Figure 3.6D-E), providing a functional link between KDM5B and ESC proliferation. Interestingly, affinity purification of KDM5B demethylase complex in ESCs pulled down proteins directly involved in cell cycle control and DNA synthesis including pre-replication complex (Pre-RC) components origin recognition complexes (ORC), minichromosome maintenance complex (MCM), and the histone chaperone FACT (Figure 3.7B-C). The ORC complex recognizes replication origins and recruits DNA helicase MCM2-7 complex and FACT components. FACT can disassemble the nucleosome barrier ahead of the replication fork and permit the MCM complex to unwind DNA (Abe et al, 2011; Gilbert, 2010). Interestingly, many of these KDM5B-interacting proteins are encoded by KDM5B targets, suggesting a positive feedback loop that reinforces robust DNA synthesis and proliferation. Similarly, the KDM5B Drosophila homologue Lid also interacts with multiple Mcm proteins (Moshkin et al, 2009), suggesting an evolutionarily conserved role for KDM5 family members in cell cycle control by directly interacting with DNA replication machinery. Pre-RC may load

many replication origins on chromatin but many binding sites are dormant. It is not clear which sites determine replication firing as well as replication timing kinetics (Gilbert, 2010). Because chromatin marked by MRG15 and H3K36me3 showed early replication timing and higher Pre-RC binding (Filion et al, 2010), it is tempting to speculate that KDM5B's association with Pre-RC may contribute to replication firing and timing. Moreover, DNA replication entails global disassembly and reassembly of nucleosomes in the wake of the DNA replication fork. The association of KDM5B with histone chaperone proteins FACT and Asf1b (Figure 3.7C) (Moshkin et al, 2009) may facilitate the resetting of histone modifications and preservation of epigenetic memory. The ability of KDM5B to directly interact with DNA replication machinery echoes its role in regulating the ESC cell cycle and high proliferative rate. ESCs show atypical cell cycle regulation characterized by high portion of S phase and short G1 phase (White & Dalton, 2005). We propose that KDM5B regulated high rate of proliferation and S phase entry can forcibly establish and maintain a self-renewal associated transcriptional program and epigenetic landscape, which is important to achieving naïve pluripotency. We are testing these hypotheses in follow up experiments.

8. Further evidence linking KDM5B to intragenic H3K36me3

Proteomics analysis of the KDM5B demethylase complex provides additional evidence linking KDM5B to transcriptional elongation. We used reciprocal Co-IP to validate these interactions (Appendix Figure 2-3). HDAC6 has been found to interact with KDM5B (Barrett et al, 2007) as well as elongating RNA Pol II (Wang et al, 2009b). mRNA export adaptor protein THOC4/Yra1/ALY has been shown to bind directly to the hyperphosphorylated CTD of RNA Pol II and to avert premature mRNA export (MacKellar & Greenleaf, 2011). SSRP1 is a component of the FACT complex required

for productive transcription elongation on chromatin templates *in vitro* (Orphanides et al, 1998). The FACT complex possesses histone chaperone activity and could promote H2A-H2B eviction in front of RNA Pol II and facilitates elongating Pol II passage (Orphanides et al, 1999).

Our identification of the H3K36 methylation writer and reader protein NSD3 as a KDM5B interacting protein provides additional evidence connecting KDM5B to transcriptional elongation. NSD3 is a SET domain containing protein that can catalyze methylation of H3K36 and contains a PWWP domain that specifically recognizes H3K36me3 (Wagner & Carpenter, 2012). Knockdown of *Nsd3* also decreases H3K36me3 *in vivo* (Rahman et al, 2011). Interestingly, NSD3 is also part of another intragenic histone demethylase complex associated with elongating Pol II and P-TEFb, further suggesting a role in transcription elongation (Fang et al, 2010). Moreover, the *Nsd3* gene is amplified in human breast cancer cell lines (Angrand et al, 2001) in which *Kdm5b* is often up-regulated (Lu et al, 1999; Yamane et al, 2007). We previously reported that MRG15 could recruit KDM5B to intragenic regions. *Kdm5b* knockout embryos are peri-implantation lethal (Catchpole et al, 2011) whereas *Mrg15* knockout mice die at later embryonic stage (~E14.5) (Tominaga et al, 2005). This suggests that additional KDM5B recruitment mechanisms exist. We examined whether NSD3 contributes to intragenic recruitment of KDM5B. First, we confirmed KDM5B-NSD3 interaction in ESCs and 293 cells. Second, our preliminary data suggest that knock down of *Nsd3* decreased KDM5B intragenic occupancy selectively at *Kdm5b* target sites. Third, *Nsd3* knockdown selectively decreased the expression of several KDM5B target genes (Appendix Figure 3). These observations suggest additional mechanisms by which KDM5B can be recruited to gene bodies and echo our previous observation

that KDM5B is an important regulator of transcriptional elongation at transcribed chromatin.

In summary, proteomics analysis of KDM5B-associated interactions reveals that KDM5B may be a multifunctional chromatin modification enzyme participating in a variety of chromatin-based nuclear processes, including transcription, DNA replication, mRNA processing (splicing, poly-adenylation, mRNA export) *et al.* We have provided multiple lines of evidence that KDM5B is targeted to actively transcribed chromatin via its tight association with H3K36me3. MRG15 and NSD3 also interact and recruit KDM5B to H3K36me3-enriched chromatin. Accumulating evidence suggests that H3K36me3 machinery can actively modulate transcription (Carrozza et al, 2005; Joshi & Struhl, 2005; Keogh et al, 2005; Lin et al, 2008; Morris et al, 2005), DNA replication (Filion et al, 2010; Kim et al, 2008a; Pryde et al, 2009), mRNA processing (Kolasinska-Zwierz et al, 2009; Luco et al, 2010; Schwartz et al, 2009; Spies et al, 2009). We propose that H3K36me3 is a major mechanism linking KDM5B to transcriptional elongation, mRNA processing and possibly DNA replication. Moreover, we propose that Nanog downstream target *Kdm5b* is an important epigenetic regulator of ESC self-renewal associated gene-network and DNA replication, which contributes to the highly proliferative state and naïve pluripotency.

Conclusion

As summarized in Figure 4.1, I propose that KDM5B is the major epigenetic module downstream of Nanog to mediate Nanog functionality in establishing and maintaining naïve pluripotent state. KDM5B uniquely activates a pluripotency-associated gene network by repressing intragenic spurious initiation and shaping a H3K4me3 gradient important for productive transcriptional elongation. KDM5B sustains high proliferative rate of ESCs by activating the cell cycle progression and DNA synthesis gene-network module and possibly by directly interacting with DNA replication machinery, in which the naïve pluripotent state is forcibly established and inherited.

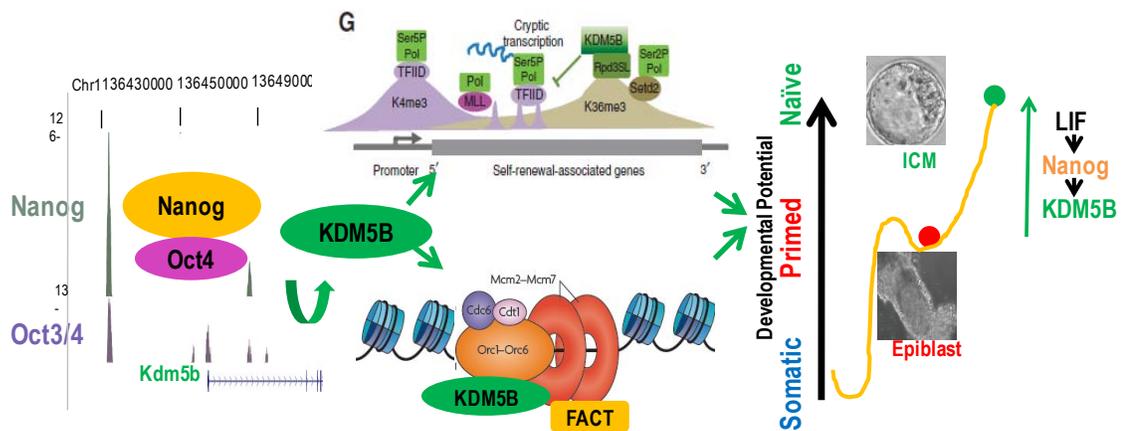
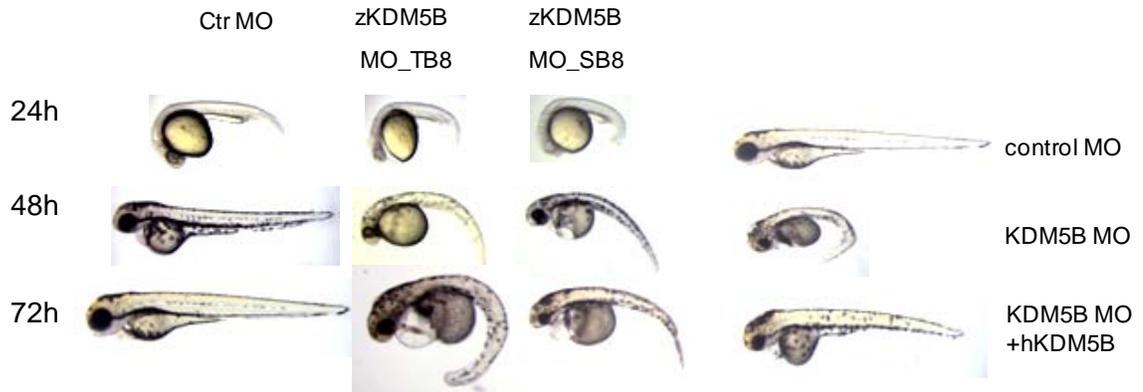


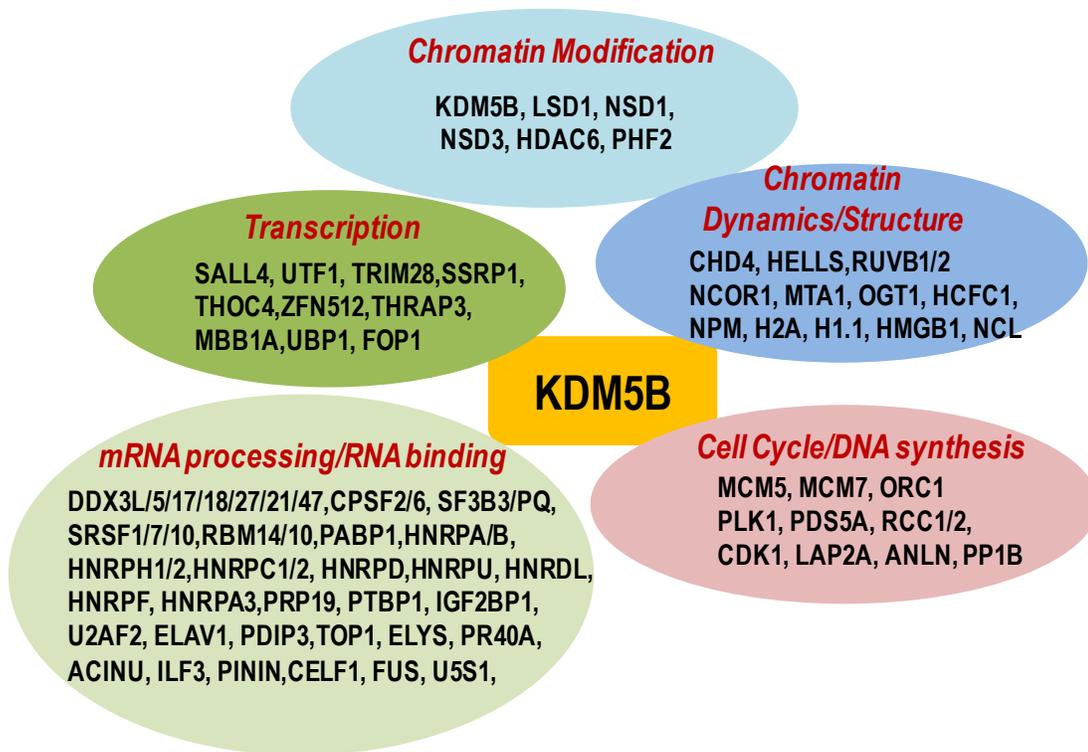
Figure 4.1 Model of KDM5B in regulating naïve pluripotent state.

APPENDICES

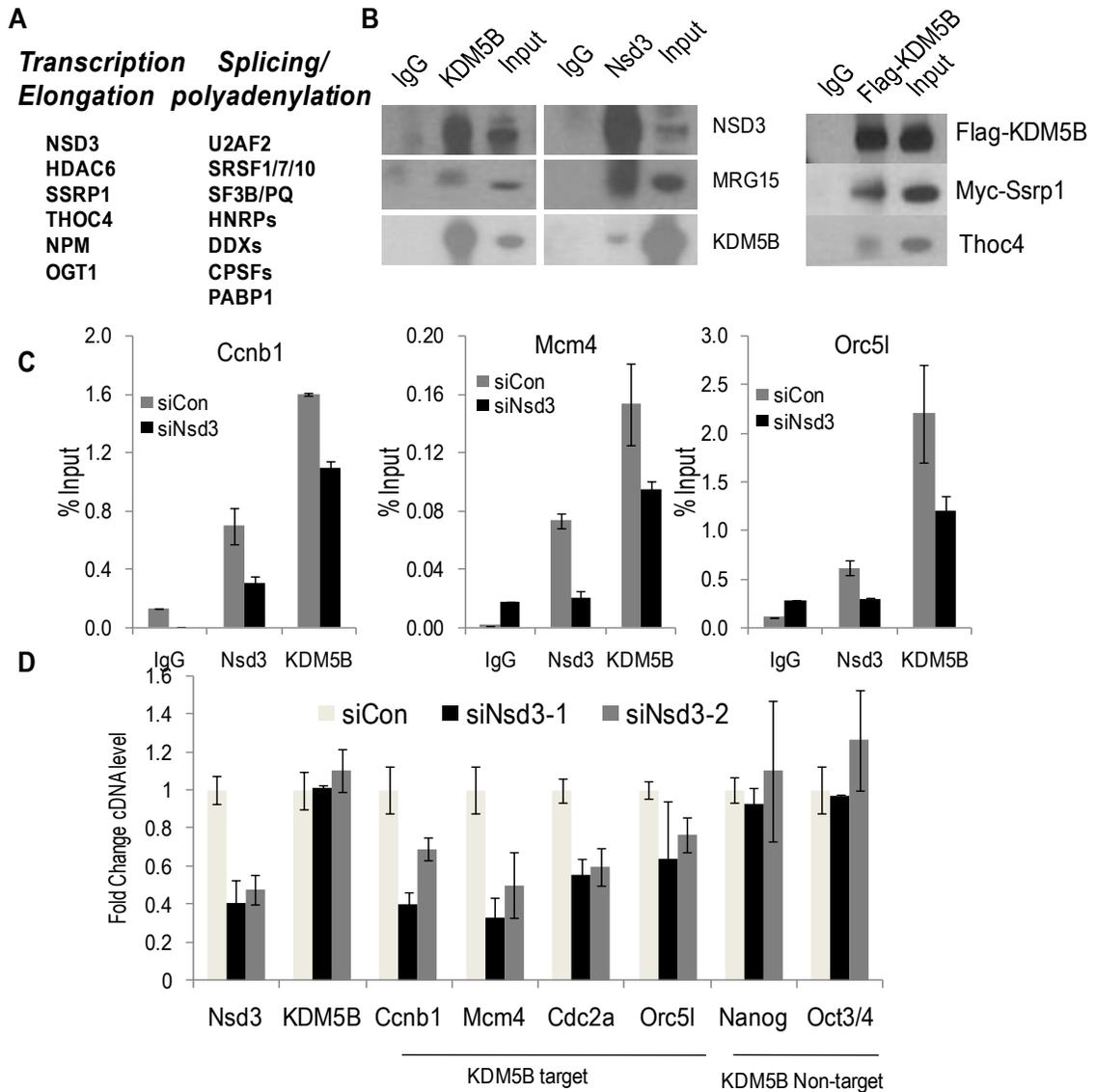
Figures



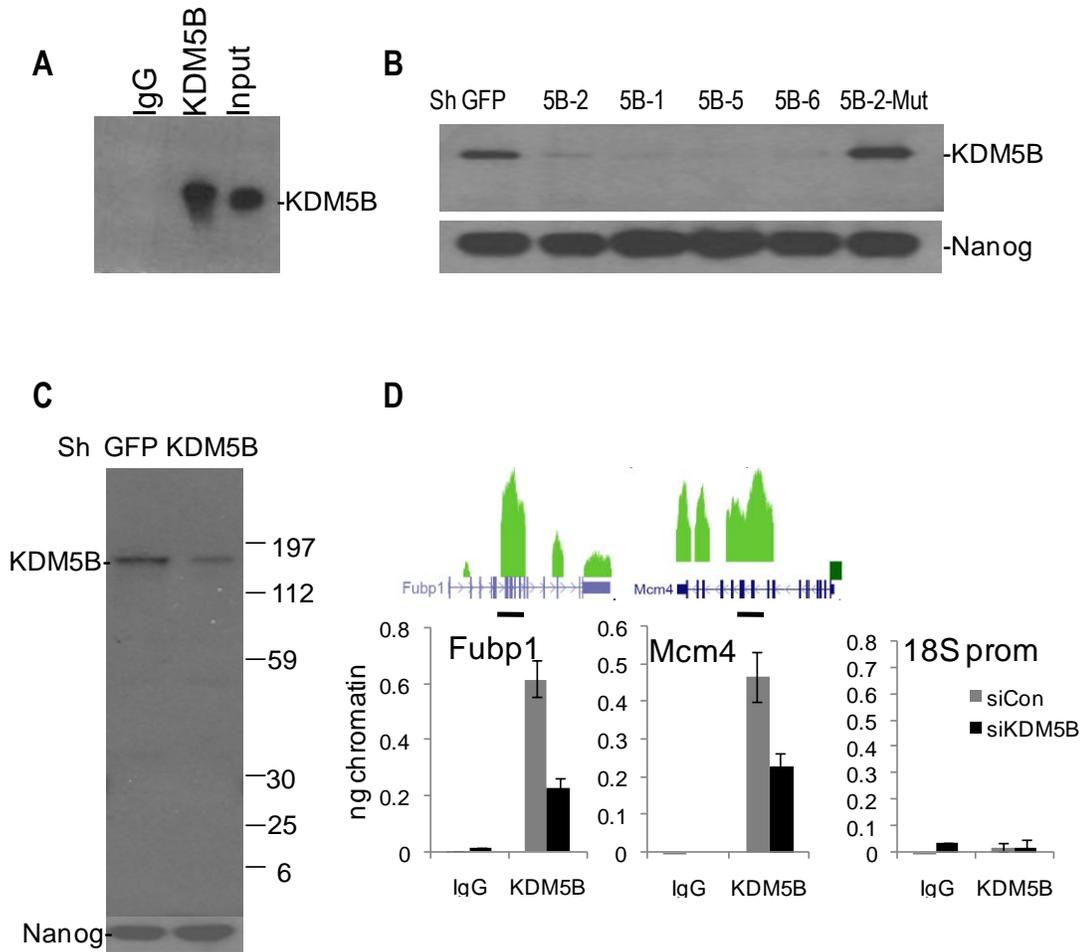
Appendix Figure 1: Inhibition of one zebrafish Kdm5b paralogue (ENSDARG0000003098) on chromosome 8 by two antisense morpholino oligonucleotides (MO)—one blocking translation (MO_TB8) and the other blocking splicing (MO_SB8). The effect of MO_SB8 was confirmed by RT-PCR that resulted in the exclusion of exon 11 and production of a 633 amino acid truncated protein lacking the catalytic JmjC domain. Similar results were obtained using a translation blocking MO targeting to another zKdm5b paralogue (ENSDARG00000057093) on chromosome 11. Injection of high doses of anti-zKdm5b MO resulted in a severe pre-gastrula developmental arrest. Injection of lower dosage of MO caused a significant developmental arrest, characterized by a smaller eye and body length, curved tail and heart edema (left). No zKdm5b MO treated fish survived beyond 3 days post-fertilization and co-injection of human Kdm5b mRNA could rescue the developmental defect (right).



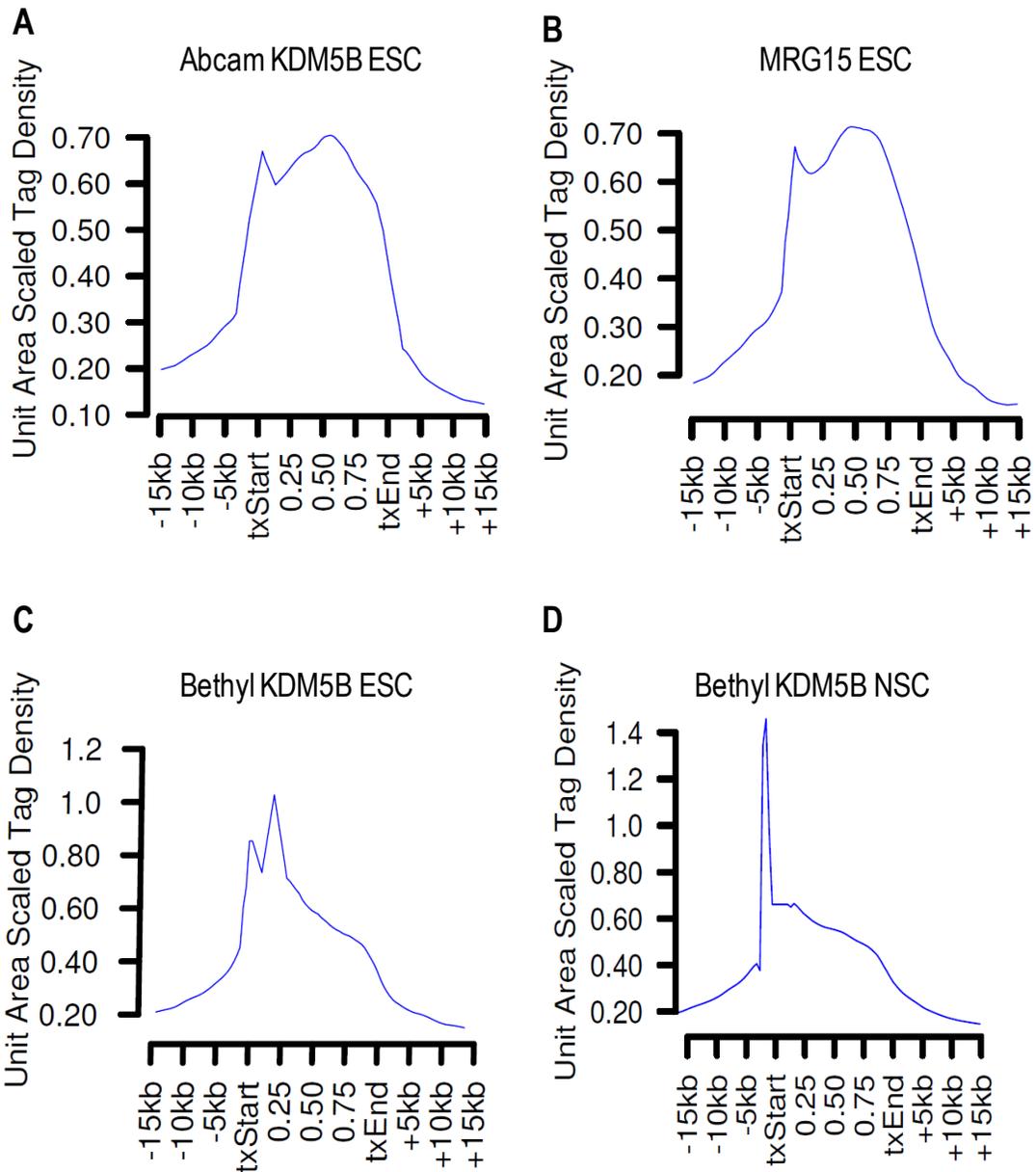
Appendix Figure 2: Functional classification and graphical visualization of components of the KDM5B demethylase complex identified by mass spectrometry proteomics analysis. KDM5B-associated proteins can be categorized into the following major groups: transcription, chromatin modification, chromatin structure and dynamics, cell cycle and DNA synthesis and mRNA processing or RNA binding.



Appendix Figure 3: (A) List of potential Kdm5b-interacting polypeptides involved in transcriptional elongation and mRNA processing. **(B)** Reciprocal Co-IP confirmation of Kdm5b interaction with Nsd3, Ssrp1 and Thoc4. **(C)** Nsd3 and Kdm5b ChIP analysis on the intragenic Kdm5b ChIP-Seq peak region after shRNA-mediated Nsd3 knockdown. **(D)** qRT-PCR analysis of the expression of Kdm5b activated target genes following Nsd3 knockdown.



Appendix Figure 4: (A) IP-WB of Kdm5b in J1 ESCs using ab50958 anti-Kdm5b antibody (affinity purified). (B) WB analysis of Kdm5b from J1 ESCs transfected with multiple Kdm5b shRNAs (5B-1,2,5,6). 5B-2-Mut indicates a shRNA point mutant that abolishes its RNAi activity. Nanog was used as loading control. (C) WB analysis of Kdm5b following Kdm5b knockdown in J1 ESCs. Nuclear extract was harvested. (D) ChIP-qPCR of Kdm5b intragenic ChIP signal following Kdm5b shRNA knockdown at indicated ChIP-Seq peak regions. 18S promoter was used as a negative control. Kdm5b ab50958 affinity purified antibody was used for all above experiments.



Appendix Figure 5: Gene scaled tag profiles of deep sequenced Kdm5b ChIP-Seq using abcam50958 antibody in ESCs **(A)**. Kdm5b ChIP-Seq was also performed using Bethyl A301-813A anti-Kdm5b antibody in ESCs **(C)** and neural stem cells **(D)**. Deep sequenced MRG15 ChIP-Seq gene scaled tag profile in ESCs was also plotted in **(B)**.

Contributions to Figures

Carl Pelz developed the pipeline to analyze high-throughput sequencing data and performed most of the bioinformatics analysis and gene-scaled ChIP-Seq visualization using R in Chapter Two.

Soren Impey did immunofluorescence staining of Kdm5b in early mouse embryos in Figure 3.2B.

Annelis Haft did the surgical injection of wide type ESCs and Kdm5b stable ESCs into the kidney capsule and subcutaneously of Fah immunodeficient mice in Figure 3.4D. Amir Bashar helped with taking representative pictures of three germ lineages from H.E. staining section.

Soren Impey contributed to the Kdm5b reprogramming data in Figure 3.5E.

REFERENCES

- Abe T, Sugimura K, Hosono Y, Takami Y, Akita M, Yoshimura A, Tada S, Nakayama T, Murofushi H, Okumura K, Takeda S, Horikoshi M, Seki M, Enomoto T (2011) The histone chaperone facilitates chromatin transcription (FACT) protein maintains normal replication fork rates. *J Biol Chem* **286**(35): 30504-30512
- Ahmed S, Palermo C, Wan S, Walworth NC (2004) A novel protein with similarities to Rb binding protein 2 compensates for loss of Chk1 function and affects histone modification in fission yeast. *Mol Cell Biol* **24**(9): 3660-3669
- Allis CD, Berger SL, Cote J, Dent S, Jenuwien T, Kouzarides T, Pillus L, Reinberg D, Shi Y, Shiekhhattar R, Shilatifard A, Workman J, Zhang Y (2007) New nomenclature for chromatin-modifying enzymes. *Cell* **131**(4): 633-636
- Ang YS, Tsai SY, Lee DF, Monk J, Su J, Ratnakumar K, Ding J, Ge Y, Darr H, Chang B, Wang J, Rendl M, Bernstein E, Schaniel C, Lemischka IR (2011) Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**(2): 183-197
- Angrand PO, Apiou F, Stewart AF, Dutrillaux B, Losson R, Chambon P (2001) NSD3, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines. *Genomics* **74**(1): 79-88
- Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17**(1): 126-140
- Bao S, Tang F, Li X, Hayashi K, Gillich A, Lao K, Surani MA (2009) Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature* **461**(7268): 1292-1295
- Barrett A, Madsen B, Copier J, Lu PJ, Cooper L, Scibetta AG, Burchell J, Taylor-Papadimitriou J (2002) PLU-1 nuclear protein, which is upregulated in breast cancer, shows restricted expression in normal human adult tissues: a new cancer/testis antigen? *Int J Cancer* **101**(6): 581-588
- Barrett A, Santangelo S, Tan K, Catchpole S, Roberts K, Spencer-Dene B, Hall D, Scibetta A, Burchell J, Verdin E, Freemont P, Taylor-Papadimitriou J (2007) Breast cancer associated transcriptional repressor PLU-1/JARID1B interacts directly with histone deacetylases. *Int J Cancer* **121**(2): 265-275
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* **129**(4): 823-837
- Belotserkovskaya R, Reinberg D (2004) Facts about FACT and transcript elongation through chromatin. *Curr Opin Genet Dev* **14**(2): 139-146

- Benevolenskaya EV, Murray HL, Branton P, Young RA, Kaelin WG, Jr. (2005) Binding of pRB to the PHD protein RBP2 promotes cellular differentiation. *Mol Cell* **18**(6): 623-635
- Bernstein BE, Meissner A, Lander ES (2007) The mammalian epigenome. *Cell* **128**(4): 669-681
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**(2): 315-326
- Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ, Klose RJ (2010) CpG islands recruit a histone H3 lysine 36 demethylase. *Mol Cell* **38**(2): 179-190
- Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**(6): 947-956
- Bradley A, Evans M, Kaufman MH, Robertson E (1984) Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**(5965): 255-256
- Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, Howlett SK, Clarkson A, Ahrlund-Richter L, Pedersen RA, Vallier L (2007) Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**(7150): 191-195
- Buecker C, Chen HH, Polo JM, Daheron L, Bu L, Barakat TS, Okwieka P, Porter A, Gribnau J, Hochedlinger K, Geijsen N (2010) A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. *Cell Stem Cell* **6**(6): 535-546
- Capecchi MR (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* **6**(6): 507-512
- Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP, Workman JL (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**(4): 581-592
- Cartwright P, McLean C, Sheppard A, Rivett D, Jones K, Dalton S (2005) LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development* **132**(5): 885-896
- Catchpole S, Spencer-Dene B, Hall D, Santangelo S, Rosewell I, Guenatri M, Beatson R, Scibetta AG, Burchell JM, Taylor-Papadimitriou J (2011) PLU-1/JARID1B/KDM5B is required for embryonic survival and contributes to cell proliferation in the mammary gland and in ER+ breast cancer cells. *Int J Oncol* **38**(5): 1267-1277
- Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, Smith A (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**(5): 643-655

Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, Robertson M, Vrana J, Jones K, Grotewold L, Smith A (2007) Nanog safeguards pluripotency and mediates germline development. *Nature* **450**(7173): 1230-1234

Chan SW, Hong W (2001) Retinoblastoma-binding protein 2 (Rbp2) potentiates nuclear hormone receptor-mediated transcription. *J Biol Chem* **276**(30): 28402-28412

Chazaud C, Yamanaka Y, Pawson T, Rossant J (2006) Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev Cell* **10**(5): 615-624

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK, Clarke ND, Wei CL, Ng HH (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**(6): 1106-1117

Chen Z, Zang J, Whetstine J, Hong X, Davrazou F, Kutateladze TG, Simpson M, Mao Q, Pan CH, Dai S, Hagman J, Hansen K, Shi Y, Zhang G (2006) Structural insights into histone demethylation by JMJD2 family members. *Cell* **125**(4): 691-702

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**(13): 3497-3500

Christensen J, Agger K, Cloos PA, Pasini D, Rose S, Sennels L, Rappsilber J, Hansen KH, Salcini AE, Helin K (2007) RBP2 belongs to a family of demethylases, specific for tri- and dimethylated lysine 4 on histone 3. *Cell* **128**(6): 1063-1076

Coue M, Kearsley SE, Mechali M (1996) Chromatin binding, nuclear localization and phosphorylation of *Xenopus cdc21* are cell-cycle dependent and associated with the control of initiation of DNA replication. *EMBO J* **15**(5): 1085-1097

Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, Davies H, Edkins S, Hardy C, Latimer C, Teague J, Andrews J, Barthorpe S, Beare D, Buck G, Campbell PJ, Forbes S, Jia M, Jones D, Knott H, Kok CY, Lau KW, Leroy C, Lin ML, McBride DJ, Maddison M, Maguire S, McLay K, Menzies A, Mironenko T, Mulderrig L, Mudie L, O'Meara S, Pleasance E, Rajasingham A, Shepherd R, Smith R, Stebbings L, Stephens P, Tang G, Tarpey PS, Turrell K, Dykema KJ, Khoo SK, Petillo D, Wonderegem B, Anema J, Kahnoski RJ, Teh BT, Stratton MR, Futreal PA (2010) Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**(7279): 360-363

de Almeida SF, Grosso AR, Koch F, Fenouil R, Carvalho S, Andrade J, Levezinho H, Gut M, Eick D, Gut I, Andrau JC, Ferrier P, Carmo-Fonseca M (2011) Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat Struct Mol Biol* **18**(9): 977-983

Dejosez M, Zwaka TP (2012) Pluripotency and Nuclear Reprogramming. *Annu Rev Biochem*

- Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**(5): P3
- Dey BK, Stalker L, Schnerch A, Bhatia M, Taylor-Papadimitriou J, Wynder C (2008) The histone demethylase KDM5b/JARID1b plays a role in cell fate decisions by blocking terminal differentiation. *Mol Cell Biol* **28**(17): 5312-5327
- Dignam JD, Lebovitz RM, Roeder RG (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **11**(5): 1475-1489
- Ding J, Xu H, Faiola F, Ma'ayan A, Wang J (2011) Oct4 links multiple epigenetic pathways to the pluripotency network. *Cell Res* **22**(1): 155-167
- Do JT, Scholer HR (2009) Regulatory circuits underlying pluripotency and reprogramming. *Trends Pharmacol Sci* **30**(6): 296-302
- Doyon Y, Selleck W, Lane WS, Tan S, Cote J (2004) Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. *Mol Cell Biol* **24**(5): 1884-1896
- Edelmann W, Yang K, Umar A, Heyer J, Lau K, Fan K, Liedtke W, Cohen PE, Kane MF, Lipford JR, Yu N, Crouse GF, Pollard JW, Kunkel T, Lipkin M, Kolodner R, Kucherlapati R (1997) Mutation in the mismatch repair gene Msh6 causes cancer susceptibility. *Cell* **91**(4): 467-477
- Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoepfner DJ, Dash C, Bazett-Jones DP, Le Grice S, McKay RD, Buetow KH, Gingeras TR, Misteli T, Meshorer E (2008) Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2**(5): 437-447
- Eissenberg JC, Lee MG, Schneider J, Ilvarsonn A, Shiekhattar R, Shilatifard A (2007) The trithorax-group gene in *Drosophila* little imaginal discs encodes a trimethylated histone H3 Lys4 demethylase. *Nat Struct Mol Biol* **14**(4): 344-346
- Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**(5819): 154-156
- Falnes PO, Johansen RF, Seeberg E (2002) AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* **419**(6903): 178-182
- Fang R, Barbera AJ, Xu Y, Rutenberg M, Leonor T, Bi Q, Lan F, Mei P, Yuan GC, Lian C, Peng J, Cheng D, Sui G, Kaiser UB, Shi Y, Shi YG (2010) Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation. *Mol Cell* **39**(2): 222-233
- Fattaey AR, Helin K, Dembski MS, Dyson N, Harlow E, Vuocolo GA, Hanobik MG, Haskell KM, Oliff A, Defeo-Jones D, et al. (1993) Characterization of the retinoblastoma binding proteins RBP1 and RBP2. *Oncogene* **8**(11): 3149-3156
- Fazio TG, Panning B (2010) Condensin complexes regulate mitotic progression and interphase chromatin structure in embryonic stem cells. *J Cell Biol* **188**(4): 491-503

- Fejes AP, Robertson G, Bilenky M, Varhol R, Bainbridge M, Jones SJ (2008) FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. *Bioinformatics* **24**(15): 1729-1730
- Filion GJ, van Bommel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ, van Steensel B (2010) Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**(2): 212-224
- Frankenberg S, Smith L, Greenfield A, Zernicka-Goetz M (2007) Novel gene expression patterns along the proximo-distal axis of the mouse embryo before gastrulation. *BMC Dev Biol* **7**: 8
- Freudenberg JM, Ghosh S, Lackford BL, Yellaboina S, Zheng X, Li R, Cuddapah S, Wade PA, Hu G, Jothi R (2012) Acute depletion of Tet1-dependent 5-hydroxymethylcytosine levels impairs LIF/Stat3 signaling and results in loss of embryonic stem cell identity. *Nucleic Acids Res*
- Gambus A, Jones RC, Sanchez-Diaz A, Kanemaki M, van Deursen F, Edmondson RD, Labib K (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol* **8**(4): 358-366
- Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M (2010) Open chromatin in pluripotency and reprogramming. *Nat Rev Mol Cell Biol* **12**(1): 36-47
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**(10): R80
- Gilbert DM (2010) Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat Rev Genet* **11**(10): 673-684
- Gildea JJ, Lopez R, Shearn A (2000) A screen for new trithorax group genes identified little imaginal discs, the *Drosophila melanogaster* homologue of human retinoblastoma binding protein 2. *Genetics* **156**(2): 645-663
- Greer EL, Maures TJ, Hauswirth AG, Green EM, Leeman DS, Maro GS, Han S, Banko MR, Gozani O, Brunet A (2010) Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in *C. elegans*. *Nature* **466**(7304): 383-387
- Greer EL, Maures TJ, Ucar D, Hauswirth AG, Mancini E, Lim JP, Benayoun BA, Shi Y, Brunet A (2011) Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans*. *Nature* **479**(7373): 365-371
- Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**(1): 77-88

- Guo G, Smith A (2010) A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. *Development* **137**(19): 3185-3192
- Guo G, Yang J, Nichols J, Hall JS, Eyres I, Mansfield W, Smith A (2009) Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**(7): 1063-1069
- Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**(7364): 295-300
- Han J, Yuan P, Yang H, Zhang J, Soh BS, Li P, Lim SL, Cao S, Tay J, Orlov YL, Lufkin T, Ng HH, Tam WL, Lim B (2010) Tbx3 improves the germ-line competency of induced pluripotent stem cells. *Nature* **463**(7284): 1096-1100
- Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* **144**(5): 646-674
- Hanna J, Markoulaki S, Mitalipova M, Cheng AW, Cassady JP, Staerk J, Carey BW, Lengner CJ, Foreman R, Love J, Gao Q, Kim J, Jaenisch R (2009) Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* **4**(6): 513-524
- Hao H, Kim DS, Klocke B, Johnson KR, Cui K, Gotoh N, Zang C, Gregorski J, Gieser L, Peng W, Fann Y, Seifert M, Zhao K, Swaroop A (2012) Transcriptional Regulation of Rod Photoreceptor Homeostasis Revealed by In Vivo NRL Targetome Analysis. *PLoS Genet* **8**(4): e1002649
- Hargreaves DC, Horng T, Medzhitov R (2009) Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* **138**(1): 129-145
- Hayakawa T, Ohtani Y, Hayakawa N, Shinmyozu K, Saito M, Ishikawa F, Nakayama J (2007) RBP2 is an MRG15 complex component and down-regulates intragenic histone H3 lysine 4 methylation. *Genes Cells* **12**(6): 811-826
- Hayami S, Yoshimatsu M, Veerakumarasivam A, Unoki M, Iwai Y, Tsunoda T, Field HI, Kelly JD, Neal DE, Yamaue H, Ponder BA, Nakamura Y, Hamamoto R (2010) Overexpression of the JmjC histone demethylase KDM5B in human carcinogenesis: involvement in the proliferation of cancer cells through the E2F/RB pathway. *Mol Cancer* **9**: 59
- Hochedlinger K, Plath K (2009) Epigenetic reprogramming and induced pluripotency. *Development* **136**(4): 509-523
- Hooper M, Hardy K, Handyside A, Hunter S, Monk M (1987) HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**(6110): 292-295
- Hu G, Kim J, Xu Q, Leng Y, Orkin SH, Elledge SJ (2009) A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev* **23**(7): 837-848

- Huang PH, Chen CH, Chou CC, Sargeant AM, Kulp SK, Teng CM, Byrd JC, Chen CS (2010) Histone deacetylase inhibitors stimulate histone H3 lysine 4 methylation in part via transcriptional repression of histone H3 lysine 4 demethylases. *Mol Pharmacol* **79**(1): 197-206
- Impey S, McCorkle SR, Cha-Molstad H, Dwyer JM, Yochum GS, Boss JM, McWeeney S, Dunn JJ, Mandel G, Goodman RH (2004) Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell* **119**(7): 1041-1054
- Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**(7310): 1129-1133
- Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, DeCoste C, Schafer X, Lun Y, Lemischka IR (2006) Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**(7102): 533-538
- Iwase S, Lan F, Bayliss P, de la Torre-Ubieta L, Huarte M, Qi HH, Whetstine JR, Bonni A, Roberts TM, Shi Y (2007) The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell* **128**(6): 1077-1088
- Jaenisch R, Young R (2008) Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**(4): 567-582
- Jiang J, Chan YS, Loh YH, Cai J, Tong GQ, Lim CA, Robson P, Zhong S, Ng HH (2008) A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat Cell Biol* **10**(3): 353-360
- Joshi AA, Struhl K (2005) Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol Cell* **20**(6): 971-978
- Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**(7314): 430-435
- Kaji K, Caballero IM, MacLeod R, Nichols J, Wilson VA, Hendrich B (2006) The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat Cell Biol* **8**(3): 285-292
- Kalmar T, Lim C, Hayward P, Munoz-Descalzo S, Nichols J, Garcia-Ojalvo J, Martinez Arias A (2009) Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol* **7**(7): e1000149
- Kee K, Pera RA, Turek PJ (2010) Testicular germline stem cells. *Nat Rev Urol* **7**(2): 94-100
- Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, Schuldiner M, Chin K, Punna T, Thompson NJ, Boone C, Emili A, Weissman JS, Hughes TR, Strahl BD, Grunstein M, Greenblatt JF, Buratowski S, Krogan NJ (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**(4): 593-605

- Kim HS, Rhee DK, Jang YK (2008a) Methylations of histone H3 lysine 9 and lysine 36 are functionally linked to DNA replication checkpoint control in fission yeast. *Biochem Biophys Res Commun* **368**(2): 419-425
- Kim JB, Zaehres H, Wu G, Gentile L, Ko K, Sebastiano V, Arauzo-Bravo MJ, Ruau D, Han DW, Zenke M, Scholer HR (2008b) Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature* **454**(7204): 646-650
- Kleinsmith LJ, Pierce GB, Jr. (1964) Multipotentiality of Single Embryonal Carcinoma Cells. *Cancer Res* **24**: 1544-1551
- Klose RJ, Kallin EM, Zhang Y (2006) JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet* **7**(9): 715-727
- Klose RJ, Yan Q, Tothova Z, Yamane K, Erdjument-Bromage H, Tempst P, Gilliland DG, Zhang Y, Kaelin WG, Jr. (2007) The retinoblastoma binding protein RBP2 is an H3K4 demethylase. *Cell* **128**(5): 889-900
- Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* **41**(3): 376-381
- Koledova Z, Kafkova LR, Calabkova L, Krystof V, Dolezel P, Divoky V (2009) Cdk2 inhibition prolongs G1 phase progression in mouse embryonic stem cells. *Stem Cells Dev* **19**(2): 181-194
- Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128**(4): 693-705
- Krogan NJ, Dover J, Wood A, Schneider J, Heidt J, Boateng MA, Dean K, Ryan OW, Golshani A, Johnston M, Greenblatt JF, Shilatifard A (2003) The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell* **11**(3): 721-729
- Kunath T, Saba-El-Leil MK, Almousailleakh M, Wray J, Meloche S, Smith A (2007) FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**(16): 2895-2902
- Landeira D, Sauer S, Poot R, Dvorkina M, Mazzarella L, Jorgensen HF, Pereira CF, Leleu M, Piccolo FM, Spivakov M, Brookes E, Pombo A, Fisher C, Skarnes WC, Snoek T, Bezstarosti K, Demmers J, Klose RJ, Casanova M, Tavares L, Brockdorff N, Merkenschlager M, Fisher AG (2010) Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat Cell Biol* **12**(6): 618-624
- Lange C, Calegari F (2010) Cdks and cyclins link G1 length and differentiation of embryonic, neural and hematopoietic stem cells. *Cell Cycle* **9**(10): 1893-1900

- Lee G, White LS, Hurov KE, Stappenbeck TS, Piwnica-Worms H (2009a) Response of small intestinal epithelial cells to acute disruption of cell division through CDC25 deletion. *Proc Natl Acad Sci U S A* **106**(12): 4701-4706
- Lee JS, Shilatifard A (2007) A site to remember: H3K36 methylation a mark for histone deacetylation. *Mutat Res* **618**(1-2): 130-134
- Lee MG, Norman J, Shilatifard A, Shiekhattar R (2007a) Physical and functional association of a trimethyl H3K4 demethylase and Ring6a/MBLR, a polycomb-like protein. *Cell* **128**(5): 877-887
- Lee N, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y (2009b) The H3K4 demethylase lid associates with and inhibits histone deacetylase Rpd3. *Mol Cell Biol* **29**(6): 1401-1410
- Lee N, Zhang J, Klose RJ, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y (2007b) The trithorax-group protein Lid is a histone H3 trimethyl-Lys4 demethylase. *Nat Struct Mol Biol* **14**(4): 341-343
- Letunic I, Doerks T, Bork P (2011) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* **40**(Database issue): D302-305
- Li B, Carey M, Workman JL (2007a) The role of chromatin during transcription. *Cell* **128**(4): 707-719
- Li B, Gogol M, Carey M, Pattenden SG, Seidel C, Workman JL (2007b) Infrequently transcribed long genes depend on the Set2/Rpd3S pathway for accurate transcription. *Genes Dev* **21**(11): 1422-1430
- Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**(6): 915-926
- Li F, Huarte M, Zaratiegui M, Vaughn MW, Shi Y, Martienssen R, Cande WZ (2008) Lid2 is required for coordinating H3K4 and H3K9 methylation of heterochromatin and euchromatin. *Cell* **135**(2): 272-283
- Li H, Collado M, Villasante A, Strati K, Ortega S, Canamero M, Blasco MA, Serrano M (2009) The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **460**(7259): 1136-1139
- Li L, Greer C, Eisenman RN, Secombe J (2010) Essential functions of the histone demethylase lid. *PLoS Genet* **6**(11): e1001221
- Li Q, Shi L, Gui B, Yu W, Wang J, Zhang D, Han X, Yao Z, Shang Y (2011) Binding of the JmjC demethylase JARID1B to LSD1/NuRD suppresses angiogenesis and metastasis in breast cancer cells by repressing chemokine CCL14. *Cancer Res*
- Liang J, Wan M, Zhang Y, Gu P, Xin H, Jung SY, Qin J, Wong J, Cooney AJ, Liu D, Songyang Z (2008) Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat Cell Biol* **10**(6): 731-739

- Liao J, Cui C, Chen S, Ren J, Chen J, Gao Y, Li H, Jia N, Cheng L, Xiao H, Xiao L (2009) Generation of induced pluripotent stem cell lines from adult rat cells. *Cell Stem Cell* **4**(1): 11-15
- Lin CH, Li B, Swanson S, Zhang Y, Florens L, Washburn MP, Abmayr SM, Workman JL (2008) Heterochromatin protein 1a stimulates histone H3 lysine 36 demethylation by the Drosophila KDM4A demethylase. *Mol Cell* **32**(5): 696-706
- Lloret-Llinares M, Carre C, Vaquero A, de Olano N, Azorin F (2008) Characterization of Drosophila melanogaster JmjC+N histone demethylases. *Nucleic Acids Res* **36**(9): 2852-2863
- Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, Wong KY, Sung KW, Lee CW, Zhao XD, Chiu KP, Lipovich L, Kuznetsov VA, Robson P, Stanton LW, Wei CL, Ruan Y, Lim B, Ng HH (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**(4): 431-440
- Loh YH, Zhang W, Chen X, George J, Ng HH (2007) Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev* **21**(20): 2545-2557
- Lopez-Bigas N, Kisiel TA, Dewaal DC, Holmes KB, Volkert TL, Gupta S, Love J, Murray HL, Young RA, Benevolenskaya EV (2008) Genome-wide analysis of the H3K4 histone demethylase RBP2 reveals a transcriptional program controlling differentiation. *Mol Cell* **31**(4): 520-530
- Lu PJ, Sundquist K, Baeckstrom D, Poulosom R, Hanby A, Meier-Ewert S, Jones T, Mitchell M, Pitha-Rowe P, Freemont P, Taylor-Papadimitriou J (1999) A novel gene (PLU-1) containing highly conserved putative DNA/chromatin binding motifs is specifically up-regulated in breast cancer. *J Biol Chem* **274**(22): 15633-15645
- Lucio-Eterovic AK, Singh MM, Gardner JE, Veerappan CS, Rice JC, Carpenter PB (2010) Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proc Natl Acad Sci U S A* **107**(39): 16952-16957
- Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T (2010) Regulation of alternative splicing by histone modifications. *Science* **327**(5968): 996-1000
- MacKellar AL, Greenleaf AL (2011) Cotranscriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II. *J Biol Chem* **286**(42): 36385-36395
- Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, Stadtfeld M, Yachechko R, Tchieu J, Jaenisch R, Plath K, Hochedlinger K (2007) Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**(1): 55-70
- Margueron R, Reinberg D (2010) Chromatin structure and the inheritance of epigenetic information. *Nat Rev Genet* **11**(4): 285-296

- Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. *Nature* **469**(7330): 343-349
- Marion RM, Strati K, Li H, Murga M, Blanco R, Ortega S, Fernandez-Capetillo O, Serrano M, Blasco MA (2009) A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460**(7259): 1149-1153
- Marks PA, Breslow R (2007) Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug. *Nat Biotechnol* **25**(1): 84-90
- Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J, Calabrese JM, Dennis LM, Volkert TL, Gupta S, Love J, Hannett N, Sharp PA, Bartel DP, Jaenisch R, Young RA (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**(3): 521-533
- Martin GR (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**(12): 7634-7638
- Martin GR, Evans MJ (1974) The morphology and growth of a pluripotent teratocarcinoma cell line and its derivatives in tissue culture. *Cell* **2**(3): 163-172
- Martin GR, Evans MJ (1975) Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc Natl Acad Sci U S A* **72**(4): 1441-1445
- Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA, Ko MS, Niwa H (2007) Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9**(6): 625-635
- Matsui Y, Zsebo K, Hogan BL (1992) Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**(5): 841-847
- Meshorer E, Misteli T (2006) Chromatin in pluripotent embryonic stem cells and differentiation. *Nat Rev Mol Cell Biol* **7**(7): 540-546
- Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, Misteli T (2006) Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* **10**(1): 105-116
- Metzger E, Wissmann M, Yin N, Muller JM, Schneider R, Peters AH, Gunther T, Buettner R, Schule R (2005) LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**(7057): 436-439
- Metzker ML (2009) Sequencing technologies - the next generation. *Nat Rev Genet* **11**(1): 31-46
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**(7153): 553-560

- Min IM, Waterfall JJ, Core LJ, Munroe RJ, Schimenti J, Lis JT (2011) Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev* **25**(7): 742-754
- Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**(5): 631-642
- Morris SA, Shibata Y, Noma K, Tsukamoto Y, Warren E, Temple B, Grewal SI, Strahl BD (2005) Histone H3 K36 methylation is associated with transcription elongation in *Schizosaccharomyces pombe*. *Eukaryot Cell* **4**(8): 1446-1454
- Morrissy AS, Griffith M, Marra MA (2011) Extensive relationship between antisense transcription and alternative splicing in the human genome. *Genome Res* **21**(8): 1203-1212
- Moshkin YM, Kan TW, Goodfellow H, Bezstarosti K, Maeda RK, Pilyugin M, Karch F, Bray SJ, Demmers JA, Verrijzer CP (2009) Histone chaperones ASF1 and NAP1 differentially modulate removal of active histone marks by LID-RPD3 complexes during NOTCH silencing. *Mol Cell* **35**(6): 782-793
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**(5881): 1344-1349
- Neganova I, Zhang X, Atkinson S, Lako M (2009) Expression and functional analysis of G1 to S regulatory components reveals an important role for CDK2 in cell cycle regulation in human embryonic stem cells. *Oncogene* **28**(1): 20-30
- Niakan KK, Han J, Pedersen RA, Simon C, Pera RA (2012) Human pre-implantation embryo development. *Development* **139**(5): 829-841
- Nichols J, Chambers I, Taga T, Smith A (2001) Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. *Development* **128**(12): 2333-2339
- Nichols J, Smith A (2009) Naive and primed pluripotent states. *Cell Stem Cell* **4**(6): 487-492
- Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**(3): 379-391
- Nicolas E, Yamada T, Cam HP, Fitzgerald PC, Kobayashi R, Grewal SI (2007) Distinct roles of HDAC complexes in promoter silencing, antisense suppression and DNA damage protection. *Nat Struct Mol Biol* **14**(5): 372-380
- Nikolov DB, Chen H, Halay ED, Usheva AA, Hisatake K, Lee DK, Roeder RG, Burley SK (1995) Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**(6545): 119-128

- Niwa H (2007a) How is pluripotency determined and maintained? *Development* **134**(4): 635-646
- Niwa H (2007b) Open conformation chromatin and pluripotency. *Genes Dev* **21**(21): 2671-2676
- Niwa H, Burdon T, Chambers I, Smith A (1998) Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* **12**(13): 2048-2060
- Niwa H, Miyazaki J, Smith AG (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**(4): 372-376
- Niwa H, Ogawa K, Shimosato D, Adachi K (2009) A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460**(7251): 118-122
- Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, Yagi R, Rossant J (2005) Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**(5): 917-929
- Ohgushi M, Matsumura M, Eiraku M, Murakami K, Aramaki T, Nishiyama A, Muguruma K, Nakano T, Suga H, Ueno M, Ishizaki T, Suemori H, Narumiya S, Niwa H, Sasai Y (2010) Molecular pathway and cell state responsible for dissociation-induced apoptosis in human pluripotent stem cells. *Cell Stem Cell* **7**(2): 225-239
- Onder TT, Kara N, Cherry A, Sinha AU, Zhu N, Bernt KM, Cahan P, Mancarci OB, Unternaehrer J, Gupta PB, Lander ES, Armstrong SA, Daley GQ (2012) Chromatin-modifying enzymes as modulators of reprogramming. *Nature*
- Orkin SH, Hochedlinger K (2011) Chromatin connections to pluripotency and cellular reprogramming. *Cell* **145**(6): 835-850
- Orphanides G, LeRoy G, Chang CH, Luse DS, Reinberg D (1998) FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* **92**(1): 105-116
- Orphanides G, Wu WH, Lane WS, Hampsey M, Reinberg D (1999) The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* **400**(6741): 284-288
- Paling NR, Wheadon H, Bone HK, Welham MJ (2004) Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling. *J Biol Chem* **279**(46): 48063-48070
- Palmieri SL, Peter W, Hess H, Scholer HR (1994) Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Dev Biol* **166**(1): 259-267
- Papayioannou VE, Gardner RL, McBurney MW, Babinet C, Evans MJ (1978) Participation of cultured teratocarcinoma cells in mouse embryogenesis. *J Embryol Exp Morphol* **44**: 93-104

Pardo M, Lang B, Yu L, Prosser H, Bradley A, Babu MM, Choudhary J (2010) An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* **6**(4): 382-395

Pasini D, Hansen KH, Christensen J, Agger K, Cloos PA, Helin K (2008) Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes Dev* **22**(10): 1345-1355

Pinskaya M, Gourvenec S, Morillon A (2009) H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. *EMBO J* **28**(12): 1697-1707

Pryde F, Jain D, Kerr A, Curley R, Mariotti FR, Vogelauer M (2009) H3 k36 methylation helps determine the timing of cdc45 association with replication origins. *PLoS One* **4**(6): e5882

Ptaszek LM, Cowan CA (2007) New tools for genome modification in human embryonic stem cells. *Cell Stem Cell* **1**(6): 600-602

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2011) The Pfam protein families database. *Nucleic Acids Res* **40**(Database issue): D290-301

Qi X, Li TG, Hao J, Hu J, Wang J, Simmons H, Miura S, Mishina Y, Zhao GQ (2004) BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. *Proc Natl Acad Sci U S A* **101**(16): 6027-6032

Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA (2010) c-Myc regulates transcriptional pause release. *Cell* **141**(3): 432-445

Rahman S, Sowa ME, Ottinger M, Smith JA, Shi Y, Harper JW, Howley PM (2011) The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol Cell Biol* **31**(13): 2641-2652

Rayasam GV, Wendling O, Angrand PO, Mark M, Niederreither K, Song L, Lerouge T, Hager GL, Chambon P, Losson R (2003) NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J* **22**(12): 3153-3163

Resnick JL, Bixler LS, Cheng L, Donovan PJ (1992) Long-term proliferation of mouse primordial germ cells in culture. *Nature* **359**(6395): 550-551

Roesch A, Fukunaga-Kalabis M, Schmidt EC, Zabierowski SE, Brafford PA, Vultur A, Basu D, Gimotty P, Vogt T, Herlyn M (2010) A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. *Cell* **141**(4): 583-594

Roguev A, Schaft D, Shevchenko A, Aasland R, Stewart AF (2003) High conservation of the Set1/Rad6 axis of histone 3 lysine 4 methylation in budding and fission yeasts. *J Biol Chem* **278**(10): 8487-8493

- Ruiz S, Panopoulos AD, Herrerias A, Bissig KD, Lutz M, Berggren WT, Verma IM, Izpisua Belmonte JC (2010) A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. *Curr Biol* **21**(1): 45-52
- Saha RN, Wissink EM, Bailey ER, Zhao M, Fargo DC, Hwang JY, Daigle KR, Fenn JD, Adelman K, Dudek SM (2011) Rapid activity-induced transcription of Arc and other IEGs relies on poised RNA polymerase II. *Nat Neurosci* **14**(7): 848-856
- Schmitz SU, Albert M, Malatesta M, Morey L, Johansen JV, Bak M, Tommerup N, Abarategui I, Helin K (2011) Jarid1b targets genes regulating development and is involved in neural differentiation. *EMBO J* **30**(22): 4586-4600
- Schnetzer MP, Handoko L, Akhtar-Zaidi B, Bartels CF, Pereira CF, Fisher AG, Adams DJ, Flicek P, Crawford GE, Laframboise T, Tesar P, Wei CL, Scacheri PC (2010) CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genet* **6**(7): e1001023
- Schwartz S, Meshorer E, Ast G (2009) Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* **16**(9): 990-995
- Secombe J, Li L, Carlos L, Eisenman RN (2007) The Trithorax group protein Lid is a trimethyl histone H3K4 demethylase required for dMyc-induced cell growth. *Genes Dev* **21**(5): 537-551
- Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, McDermott U, Azizian N, Zou L, Fischbach MA, Wong KK, Brandstetter K, Wittner B, Ramaswamy S, Classon M, Settleman J (2010) A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**(1): 69-80
- Shi Y, Lan F, Matson C, Mulligan P, Whetstone JR, Cole PA, Casero RA (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**(7): 941-953
- Shima N, Alcaraz A, Liachko I, Buske TR, Andrews CA, Munroe RJ, Hartford SA, Tye BK, Schimenti JC (2007) A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. *Nat Genet* **39**(1): 93-98
- Silva J, Chambers I, Pollard S, Smith A (2006) Nanog promotes transfer of pluripotency after cell fusion. *Nature* **441**(7096): 997-1001
- Silva J, Nichols J, Theunissen TW, Guo G, van Oosten AL, Barrandon O, Wray J, Yamanaka S, Chambers I, Smith A (2009) Nanog is the gateway to the pluripotent ground state. *Cell* **138**(4): 722-737
- Singh AM, Dalton S (2009) The cell cycle and Myc intersect with mechanisms that regulate pluripotency and reprogramming. *Cell Stem Cell* **5**(2): 141-149
- Singhal N, Graumann J, Wu G, Arauzo-Bravo MJ, Han DW, Greber B, Gentile L, Mann M, Scholer HR (2010) Chromatin-Remodeling Components of the BAF Complex Facilitate Reprogramming. *Cell* **141**(6): 943-955

- Smith AG, Heath JK, Donaldson DD, Wong GG, Moreau J, Stahl M, Rogers D (1988) Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**(6200): 688-690
- Smith JA, White EA, Sowa ME, Powell ML, Ottinger M, Harper JW, Howley PM (2010) Genome-wide siRNA screen identifies SMCX, EP400, and Brd4 as E2-dependent regulators of human papillomavirus oncogene expression. *Proc Natl Acad Sci U S A* **107**(8): 3752-3757
- Solter D (2006) From teratocarcinomas to embryonic stem cells and beyond: a history of embryonic stem cell research. *Nat Rev Genet* **7**(4): 319-327
- Spies N, Nielsen CB, Padgett RA, Burge CB (2009) Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell* **36**(2): 245-254
- Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Plath K (2009) Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**(2): 364-377
- Stevens LC (1967) Origin of testicular teratomas from primordial germ cells in mice. *J Natl Cancer Inst* **38**(4): 549-552
- Stevens LC (1970) The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos. *Dev Biol* **21**(3): 364-382
- Stevens LC, Little CC (1954) Spontaneous Testicular Teratomas in an Inbred Strain of Mice. *Proc Natl Acad Sci U S A* **40**(11): 1080-1087
- Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**(16): 9440-9445
- Sural TH, Peng S, Li B, Workman JL, Park PJ, Kuroda MI (2008) The MSL3 chromodomain directs a key targeting step for dosage compensation of the *Drosophila melanogaster* X chromosome. *Nat Struct Mol Biol* **15**(12): 1318-1325
- Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, Ohki M, Fukuda M, Takeda N, Niida H, Kato H, Shinkai Y (2002) G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev* **16**(14): 1779-1791
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**(5929): 930-935
- Tahiliani M, Mei P, Fang R, Leonor T, Rutenberg M, Shimizu F, Li J, Rao A, Shi Y (2007) The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. *Nature* **447**(7144): 601-605

- Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**(5): 861-872
- Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**(4): 663-676
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**(10): 2731-2739
- Tan BC, Chien CT, Hirose S, Lee SC (2006) Functional cooperation between FACT and MCM helicase facilitates initiation of chromatin DNA replication. *EMBO J* **25**(17): 3975-3985
- Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**(7150): 196-199
- Theunissen TW, Silva JC (2011) Switching on pluripotency: a perspective on the biological requirement of Nanog. *Philos Trans R Soc Lond B Biol Sci* **366**(1575): 2222-2229
- Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM (1998) Embryonic stem cell lines derived from human blastocysts. *Science* **282**(5391): 1145-1147
- Tominaga K, Kirtane B, Jackson JG, Ikeno Y, Ikeda T, Hawks C, Smith JR, Matzuk MM, Pereira-Smith OM (2005) MRG15 regulates embryonic development and cell proliferation. *Mol Cell Biol* **25**(8): 2924-2937
- Trewick SC, Henshaw TF, Hausinger RP, Lindahl T, Sedgwick B (2002) Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* **419**(6903): 174-178
- Trewick SC, McLaughlin PJ, Allshire RC (2005) Methylation: lost in hydroxylation? *EMBO Rep* **6**(4): 315-320
- Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, Zhang Y (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**(7078): 811-816
- Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C, Shimotohno K, Ishikawa F, Li E, Ueda HR, Nakayama J, Okano M (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**(7): 805-814
- Tzschach A, Lenzner S, Moser B, Reinhardt R, Chelly J, Fryns JP, Kleefstra T, Raynaud M, Turner G, Ropers HH, Kuss A, Jensen LR (2006) Novel JARID1C/SMCX mutations in patients with X-linked mental retardation. *Hum Mutat* **27**(4): 389

- Usheva A, Maldonado E, Goldring A, Lu H, Houbavi C, Reinberg D, Aloni Y (1992) Specific interaction between the nonphosphorylated form of RNA polymerase II and the TATA-binding protein. *Cell* **69**(5): 871-881
- van den Berg DL, Snoek T, Mullin NP, Yates A, Bezstarosti K, Demmers J, Chambers I, Poot RA (2010) An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6**(4): 369-381
- van Zutven LJ, Onen E, Velthuisen SC, van Drunen E, von Bergh AR, van den Heuvel-Eibrink MM, Veronese A, Mecucci C, Negrini M, de Greef GE, Beverloo HB (2006) Identification of NUP98 abnormalities in acute leukemia: JARID1A (12p13) as a new partner gene. *Genes Chromosomes Cancer* **45**(5): 437-446
- Varlakhanova NV, Cotterman RF, deVries WN, Morgan J, Donahue LR, Murray S, Knowles BB, Knoepfler PS (2010) myc maintains embryonic stem cell pluripotency and self-renewal. *Differentiation* **80**(1): 9-19
- Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, Lee KK, Olsen JV, Hyman AA, Stunnenberg HG, Mann M (2010) Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**(6): 967-980
- Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HT (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**(1): 58-69
- Vezzoli A, Bonadies N, Allen MD, Freund SM, Santiveri CM, Kvinlaug BT, Huntly BJ, Gottgens B, Bycroft M (2010) Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nat Struct Mol Biol* **17**(5): 617-619
- Wagner EJ, Carpenter PB (2012) Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol* **13**(2): 115-126
- Walker E, Ohishi M, Davey RE, Zhang W, Cassar PA, Tanaka TS, Der SD, Morris Q, Hughes TR, Zandstra PW, Stanford WL (2007) Prediction and testing of novel transcriptional networks regulating embryonic stem cell self-renewal and commitment. *Cell Stem Cell* **1**(1): 71-86
- Wang GG, Cai L, Pasillas MP, Kamps MP (2007) NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nat Cell Biol* **9**(7): 804-812
- Wang GG, Song J, Wang Z, Dormann HL, Casadio F, Li H, Luo JL, Patel DJ, Allis CD (2009a) Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **459**(7248): 847-851
- Wang J, Rao S, Chu J, Shen X, Lévassieur DN, Theunissen TW, Orkin SH (2006) A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**(7117): 364-368

- Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W, Zhao K (2009b) Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138**(5): 1019-1031
- Webby CJ, Wolf A, Gromak N, Dreger M, Kramer H, Kessler B, Nielsen ML, Schmitz C, Butler DS, Yates JR, 3rd, Delahunty CM, Hahn P, Lengeling A, Mann M, Proudfoot NJ, Schofield CJ, Bottger A (2009) Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science* **325**(5936): 90-93
- Whetstine JR, Nottke A, Lan F, Huarte M, Smolikov S, Chen Z, Spooner E, Li E, Zhang G, Colaiacovo M, Shi Y (2006) Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**(3): 467-481
- White J, Dalton S (2005) Cell cycle control of embryonic stem cells. *Stem Cell Rev* **1**(2): 131-138
- Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP, Wagner EF, Metcalf D, Nicola NA, Gough NM (1988) Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**(6200): 684-687
- Workman JL (2006) Nucleosome displacement in transcription. *Genes Dev* **20**(15): 2009-2017
- Xiang Y, Zhu Z, Han G, Ye X, Xu B, Peng Z, Ma Y, Yu Y, Lin H, Chen AP, Chen CD (2007) JARID1B is a histone H3 lysine 4 demethylase up-regulated in prostate cancer. *Proc Natl Acad Sci U S A* **104**(49): 19226-19231
- Xie L, Pelz C, Wang W, Bashar A, Varlamova O, Shadle S, Impey S (2011) KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. *EMBO J* **30**(8): 1473-1484
- Yamaguchi S, Kimura H, Tada M, Nakatsuji N, Tada T (2005) Nanog expression in mouse germ cell development. *Gene Expr Patterns* **5**(5): 639-646
- Yamane K, Tateishi K, Klose RJ, Fang J, Fabrizio LA, Erdjument-Bromage H, Taylor-Papadimitriou J, Tempst P, Zhang Y (2007) PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Mol Cell* **25**(6): 801-812
- Yan Z, Wang Z, Sharova L, Sharov AA, Ling C, Piao Y, Aiba K, Matoba R, Wang W, Ko MS (2008) BAF250B-associated SWI/SNF chromatin-remodeling complex is required to maintain undifferentiated mouse embryonic stem cells. *Stem Cells* **26**(5): 1155-1165
- Yang J, van Oosten AL, Theunissen TW, Guo G, Silva JC, Smith A (2010) Stat3 activation is limiting for reprogramming to ground state pluripotency. *Cell Stem Cell* **7**(3): 319-328
- Yang P, Wang Y, Chen J, Li H, Kang L, Zhang Y, Chen S, Zhu B, Gao S (2011) RCOR2 is a subunit of the LSD1 complex that regulates ESC property and substitutes for SOX2 in reprogramming somatic cells to pluripotency. *Stem Cells* **29**(5): 791-801

- Yeom YI, Fuhrmann G, Ovitt CE, Brehm A, Ohbo K, Gross M, Hubner K, Scholer HR (1996) Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**(3): 881-894
- Ying QL, Nichols J, Chambers I, Smith A (2003) BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**(3): 281-292
- Young RA (2011) Control of the embryonic stem cell state. *Cell* **144**(6): 940-954
- Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin, II, Thomson JA (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**(5858): 1917-1920
- Zhang P, Du J, Sun B, Dong X, Xu G, Zhou J, Huang Q, Liu Q, Hao Q, Ding J (2006) Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic Acids Res* **34**(22): 6621-6628
- Zhang W, Zhang X, Liu H, Chen J, Ren Y, Huang D, Zou X, Xiao W (2010) Cdk1 is required for the self-renewal of mouse embryonic stem cells. *J Cell Biochem*
- Zhu D, Fang J, Li Y, Zhang J (2009) Mbd3, a component of NuRD/Mi-2 complex, helps maintain pluripotency of mouse embryonic stem cells by repressing trophectoderm differentiation. *PLoS One* **4**(11): e7684
- Zwaka TP, Thomson JA (2005) A germ cell origin of embryonic stem cells? *Development* **132**(2): 227-233