

**IMPROVE INDUCED PLURIPOTENT STEM CELL GENERATION
BY MANIPULATING EPIGENETIC STATUSES**

Gaoyang Liang

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the
Department of Biochemistry and Biophysics.

Chapel Hill
2012

Approved by

Yi Zhang, Ph.D.

William Marzluff, Ph.D.

Larysa Pevny, Ph.D.

Lishan Su, Ph.D.

Yue Xiong, Ph.D.

©2012
Gaoyang Liang
ALL RIGHTS RESERVED

ABSTRACT

GAOYANG LIANG: Improve induced pluripotent stem cell generation
by manipulating epigenetic statuses
(Under the direction of Dr. Yi Zhang)

Reprogramming of somatic cells to a pluripotent state can be achieved by introduction of defined transcription factors. The derived induced pluripotent stem (iPS) cells have molecular profiles and developmental potentials similar to embryonic stem (ES) cells. However, this reprogramming process is inefficient and its underlying mechanisms are poorly understood. To improve the efficiency of iPS cell generation and shed light on its mechanisms, I aimed to identify epigenetic modulations that can enhance iPS cell generation. By studying chemicals modulating epigenetic status and ES-cell enriched epigenetic factors, I demonstrate that butyrate, a histone deacetylase (HDAC) inhibitor, and Kdm2b, a histone demethylase specific for H3 lysine 36 dimethylation (H3K36me₂) are capable of facilitating iPS cell generation. Butyrate not only enhances the efficiency of iPS cell generation, but also suppresses the formation of partially reprogrammed cells and transformed cells. The enhancing effect of butyrate on reprogramming appears to depend on c-Myc and occurs early in reprogramming. Genome-wide microarray analysis shows that a set of ES cell-enriched genes are upregulated upon butyrate treatment. Kdm2b promotes iPS cell generation via its demethylase and DNA binding activities. The Kdm2b-mediated effect on reprogramming is independent of its role in suppressing senescence. Kdm2b functions at the beginning of reprogramming and enhances activation of early responsive genes in reprogramming. Kdm2b regulates gene activation by directly

binding to and demethylating its target loci. Collectively, the research in this dissertation show that iPS cell generation can be improved by manipulating epigenetic statuses, highlighting the importance of epigenetic modifications in the establishment of pluripotency.

*Dedicated to my mother,
who fosters me with endless love and care
from the beginning of my life.*

謹以此文獻吾母。

啟吾生，

澤親之愛慈無盡。

ACKNOWLEDGEMENT

I must take this opportunity to acknowledge the significant role of the following people in shaping my Ph.D. research. First and foremost, I would like to thank Dr. Yi Zhang for not only his scientific guidance and support but also his confidence in me pursuing the projects I am interested in. The Ph.D. training under his mentorship renders me invaluable experience, which becomes part of me as a scientist.

I would also like to thank my committee members, Drs. William Marzluff, Larysa Pevny, Lishan Su and Yue Xiong. Their encouragement and scientific advice continue to help advance my research.

To the members of the Zhang lab, past and present, thank you all for the wonderful friendship, insightful discussion and helpful suggestion. I am fortunate that I have worked with such a group of talented scientists. Special thanks to Drs Jin He, Olena Taranova and Jia Fang for sharing their outstanding expertise with me.

Thanks to all the friends I met in Chapel Hill. Without you along the years, my life would be monochromic. Sincere thanks to Ningqi Hou, Yu Lei, Shen Shen, Zhanpeng Huang, Jiabin Chen, Liying Zhang, Shuntai Zhou and Jianming Zhang for sharing many cheerful and unforgettable moments. Finally, I would like to conclude with my deepest thanks to my parents for their understanding, support and unconditional love.

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS.....	xiv
CHAPTER 1 INTRODUCTION	1
1.1 Epigenetics.....	1
1.2 Molecular basics for epigenetic regulations	2
DNA methylation and CpG islands	2
Histone modifications	4
Histone variants	5
Chromatin remodeling	6
Noncoding RNA	7
1.3 Genetic and epigenetic regulation of embryonic stem cells	8
Transcription factor circuitry for pluripotency	9
Epigenetic regulation of embryonic stem cells.....	10
1.4 Generation of induced pluripotent stem (iPS) cells	13
Somatic cell reprogramming to pluripotency	13
Transcription factor-directed iPS cell generation	14
Mechanisms of iPS cell generation.....	15
Models for the iPS cell generation process	15

Generation of iPS cells is a stepwise process.....	16
Roles of the introduced transcription factors	17
Additional factors that affect iPS cell generation.....	18
1.5 Introduction to the research within the dissertation.....	20
CHAPTER 2 BUTYRATE PROMOTES INDUCED PLURIPOTENT STEM CELL GENERATION	22
2.1 Introduction.....	22
2.2 Results.....	23
Butyrate promotes iPS cell generation.....	23
Butyrate accelerates iPS cell generation and its effect is c-Myc dependent	25
Sox2-GFP positive colonies are pluripotent	26
Butyrate facilitates iPS cell generation in an early time window	26
Butyrate up-regulates a set of ES cell-enriched genes in c-Myc- mediated reprogramming	27
2.3 Discussion.....	29
2.4 Materials and methods	32
MEF derivation and iPS cell culture.....	32
Retrovirus preparation and infection	33
Generation of iPS cells and calculation of reprogramming efficiency	33
Teratoma formation and analysis.....	34
Quantitative and semi-quantitative RT-PCR	34

Genome-wide expression analysis	34
Western blotting.....	35
CHAPTER 3 KDM2B PROMOTES INDUCED PLURIPOTENT STEM CELL GENERATION BY FACILITATING GENE ACTIVATION EARLY IN REPROGRAMMING	47
3.1 Introduction.....	47
3.2 Results.....	49
Kdm2b promotes iPS cell generation	49
iPS cells generated in the presence of Kdm2b are pluripotent	50
Kdm2b facilitates iPS generation in an enzymatic activity-dependent and cell proliferation alteration-independent manner	51
Kdm2b functions early during the reprogramming process	54
Kdm2b amplifies early gene activation during reprogramming	54
Kdm2b acts in concert with the key reprogramming factors in early gene activation	59
Kdm2b binds to and regulates the H3K36me2 level of the promoter of early activated genes	60
Enhanced early gene activation mediates the effect of Kdm2b in reprogramming.....	61
3.3 Discussion.....	62
Kdm2b promotes reprogramming independent of its effect on cell proliferation.....	62
A transcription cascade model for gene regulation in reprogramming.....	63
Kdm2b facilitates early gene activation in reprogramming.....	64
3.4 Materials and methods	66

Plasmids and virus preparation	66
MEF derivation and iPS cell generation	66
Cell staining, teratoma assay and chimera generation	67
RT-PCR and Western blotting	68
Microarray analysis.....	68
Chromatin Immunoprecipitation (ChIP).....	69
Standard Error Reporting	69
REFERENCE	96

LIST OF TABLES

Table S2-1	Probes of ES cell-enriched genes that are 2-fold up-regulated by butyrate in 4F reprogramming, 3F reprogramming and both.....	44
Table S2-2	Primers used for quantitative and semi-quantitative RT-PCR in Chapter Two.....	46
Table S3-1	Target sequences of shRNAs used in Chapter Three.....	91
Table S3-2	Sequences of primers used in quantitative and semi-quantitative RT-PCR in Chapter Three.....	92
Table S3-3	Sequences of primers used in ChIP-qPCR in Chapter Three.....	95

LIST OF FIGURES

Figure 2-1	Butyrate promotes iPS cells generation.	36
Figure 2-2	Butyrate improves the reprogramming kinetics in a c-Myc- dependent manner.	38
Figure 2-3	iPS cells generated in the presence of butyrate are pluripotent.....	39
Figure 2-4	Butyrate facilitates iPS cell generation at an early time window during reprogramming.....	40
Figure 2-5	Butyrate enhances the expression of a set of ES cell-enriched genes in a c-Myc-dependent manner.....	41
Figure S2-1	Expression of butyrate up-regulated genes in response to different concentration of butyrate.....	42
Figure S2-2	Butyrate treatment does not significantly suppress the p53-p21 pathway or Ink4a/Arf level.	43
Figure 3-1	Kdm2b promotes iPS cell generation.....	70
Figure 3-2	iPS cells generated by enforced expression of OSK plus Kdm2b are pluripotent.	72
Figure 3-3	Kdm2b facilitates iPS cell generation in a JmjC and ZF domain- dependent manner.	74
Figure 3-4	Kdm2b promotes iPS cell generation independently of Ink4a/Arf.	76
Figure 3-5	Kdm2b exerts its effect from the beginning of reprogramming process and facilitates gene activation in early stage of reprogramming.	77
Figure 3-6	Kdm2b directly activates early responsive genes in reprogramming.....	79
Figure 3-7	Kdm2b localizes on and demethylates the promoter region of early activated genes.	81

Figure 3-8	Inhibition of the expression of early-activated genes compromises the capacity of Kdm2b to enhance iPS cell generation.	83
Figure S3-1	Diagram of domain structure for the Kdm2 family members.	85
Figure S3-2	Induction of Kdm2b in reprogramming and ectopic expression of Kdm2b.....	86
Figure S3-3	Overexpression of Kdm2b does not affect expression of retroviral reprogramming factors.	87
Figure S3-4	Kdm2b enhances transcription activation in the examined time window.	88
Figure S3-5	RT-qPCR showing the expression of mesenchymal genes and endogenous pluripotent genes in Kdm2b-assisted reprogramming.	90

LIST OF ABBREVIATIONS

AP	alkaline phosphatase
Arf	alternative reading frame
BAF	Brg/Brahma-associated factor
Bex1	brain expressed gene 1
BMP	bone morphogenetic protein
BSA	bovine serum albumin
Cdh1	cadherin 1
Cdkn1a	cyclin-dependent kinase inhibitor 1A
cDNA	complementary DNA
CGI	CpG island
CHD	chromodomain helicase DNA binding protein
ChIP	chromatin immunoprecipitation
Cldn3/4/7	claudin 3/4/7
c-Myc	cellular myelocytomatosis oncogene
Crb3	crumbs homolog 3 (Drosophila)
cRNA	complementary RNA
DMEM	Dulbecco's Modified Eagle Medium
DNMT	DNA methyltransferase
Dox	doxycyclin
Dppa5	developmental pluripotency associated 5
Dsg2	desmoglein 2
Dsp	desmoplakin
EMT	epithelial-mesenchymal transition
Epcam	epithelial cell adhesion molecule

ES	embryonic stem
esBAF	ES cell-specific BAF complex
Esrp1	epithelial splicing regulatory protein 1
Fbox	F-box domain
FBS	fetal bovine serum
Fbx15	F-box and leucine-rich repeat protein 15
Gapdh	glyceraldehyde-3-phosphate dehydrogenase
GFP	green fluorescent protein
H3K27me3	histone H3 lysine 27 trimethylation
H3K36me2	histone H3 lysine 36 dimethylation
H3K4me2	histone H3 lysine 4 dimethylation
H3K4me3	histone H3 lysine 4 trimethylation
H3K9me3	histone H3 lysine 9 trimethylation
HDAC	histone deacetylase
HP1	heterochromatin protein 1
ICM	inner cell mass
IF	isoform
IgG	immunoglobulin G
Ink4a/b	inhibitor of Cdk4 A/B
iPS	induced pluripotent stem
ISWI	imitation switch (Drosophila)
JmjC	jumonji C
K	lysine
Kdm2b	lysine demethylase 2B
Krt12	keratin 12
LRR	leucine-rich repeats

MBD	methyl-binding domain
me(1/2/3)	(mono/di/tri-) methylation
MEF	mouse embryonic fibroblast
MES	4-morpholineethanesulfonic acid
MET	mesenchymal-to-epithelial transition
miR	microRNA
Mreg	melanoregulin
ncRNA	non-coding RNA
NEAA	non-essential amino acid
Ocln	occludin
Oct4	octamer-binding transcription factor 4
ORF	open reading frame
OSK	Oct4, Sox2 and Klf4
OSKM	Oct4, Sox2, Klf4 and c-Myc
PCR	polymerase chain reaction
PHD	plant homeo domain
PRC	polycomb repressive complex
qPCR	quantitative PCR
RT	reverse transcription
rtTA	reverse tetracycline transactivator
shRNA	small hairpin RNA
Snai1/2	snail homolog (<i>Drosophila</i>) 1/2
Sox2	SRY-box containing gene 2
SSEA-1	stage-specific embryonic antigen 1
SWI/SNF	switch/sucrose non-fermentable (yeast)
TdGF	teratocarcinoma-derived growth factor 1

Tgf- β	transforming growth factor β
Trx	trithorax
Utf1	undifferentiated embryonic cell transcription factor 1
VPA	valporic acid
WT	wild-type
Zeb1/2	zinc finger E-box binding homeobox 1/2
ZF	zinc finger

CHAPTER 1 INTRODUCTION

1.1 Epigenetics

During development of higher eukaryotes, a single totipotent zygote gives rise to all different cell types in the organism. Nearly all the cell types share the same genetic content, the genomic DNA, with the totipotent zygote, despite their distinct phenotypes and gene expression patterns. This indicates that genetic code per se is insufficient for explaining the developmental changes. How such heritable changes occur is the subject of epigenetics.

The term “epigenetics” was first coined by Conrad Waddington as “the branch of biology which studies the causal interactions between gene and their products, which bring the phenotype into being” (Waddington 1942). As our understanding of biology evolves, epigenetics is more specifically defined as the study of heritable changes in molecular or cellular phenotype that occurs without alternation in DNA sequence (Goldberg et al. 2007). Research over the past decades began to reveal the molecular mechanisms contributing to the kaleidoscopic epigenetic phenomena. These mechanisms appear to converge on the modulation of chromatin, the complex of DNA and its associated proteins and RNAs in nucleus (Bernstein et al. 2007; Martin and Zhang 2007).

1.2 Molecular basics for epigenetic regulations

In eukaryotic cells, DNA is highly condensed and packaged as chromatin in nucleus. The basic building unit of chromatin is nucleosome, in which 147 base pairs of DNA are wrapped around a histone octamer containing two H2A-H2B dimers and one (H3-H4)₂ tetramer (Luger et al. 1997; Kornberg and Lorch 1999). Nucleosomes are further assembled into higher-order structures, known as chromatin fibers, which in turn constitute a large-scale configuration known as chromosome (Woodcock 2006; Tremethick 2007). Biological processes based on DNA, such as transcription, replication and repair, must cope with the natural barrier posed by the hierarchical organization of chromatin (Misteli 2007). Chromatin regulation by covalent modifications of DNA and histones, incorporation of variant histones, nucleosome remodeling and noncoding RNAs play important roles in establishing chromatin structure and regulating DNA-based processes. These regulation strategies ultimately contribute to cellular epigenetic landscape.

DNA methylation and CpG islands

In vertebrates, genomic DNA is predominantly methylated at cytosine in the context of CpG dinucleotides with the exception of regions called CpG islands (Suzuki and Bird 2008; Deaton and Bird 2011). DNA methylation is de novo established by DNA methyltransferases (DNMTs) 3a and 3b, and maintained through replication by DNMT1, which preferentially acts on hemi-methylated DNA (Goll and Bestor 2005; Klose and Bird 2006). Deficiency in the DNA methylation machinery leads to early lethality, suggesting an essential role of DNA methylation in development (Li et al. 1992; Okano

et al. 1999). DNA methylation traditionally correlates with transcription repression. DNA methylation has been shown to fortify silencing on imprinted gene loci (Edwards and Ferguson-Smith 2007) and inactivated X chromosome (Payer and Lee 2008). Mechanistically, cytosine methylation may suppress gene expression by directly inhibiting the binding of transcription factors to DNA; alternatively, methylation can recruit methyl-binding domain (MBD) containing transcription corepressor complexes, which is capable of modifying chromatin and enforcing repression (Klose and Bird 2006).

Although once thought as a static epigenetic mark, DNA methylation is subjected to dynamic regulation. Methylation can not only be passively diluted in mitosis when Dnmt1 is inhibited or absent, but also be actively removed by repair-based mechanisms and oxidative conversion (Wu and Zhang 2010). Although the functional significance of various demethylation pathways remain to be explored, preliminary evidences suggest that oxidative conversion by Tet proteins likely maintain the methylation-free status on CpG islands, which is a cohort for transcription regulation (Ficz et al. 2011; Wu et al. 2011a; Wu et al. 2011b).

CpG islands (CGIs) are GC-rich, CpG-rich genomic DNA stretches with the length of approximate 1000 base pairs. Compared to average genomic DNA, CGIs lack DNA methylation and their unmethylated state largely persists throughout development (Suzuki and Bird 2008; Deaton and Bird 2011). Most CGIs encompass transcription initiation sites and functionally act as promoter, indicating a role in transcription regulation for these regions (Illingworth et al. 2010). Despite their unmethylated state, genes under the control of CGI promoters do not show correlation with gene expression (Suzuki and Bird 2008). However, CGI promoters show permissive signs for

transcription, as RNA polymerase II is readily detected on CGI promoters (Guenther et al. 2007; Hargreaves et al. 2009) and their local chromatin structure displays features of active transcription, including destabilized nucleosomes (Ramirez-Carrozzi et al. 2009) and enriched histone H3 lysine 4 trimethylation (H3K4me3) (Thomson et al. 2010). Silencing of CGI promoters by de novo DNA methylation and/or polycomb-mediated H3K27me3 has been thought to be important for epigenetic regulation in development (Fouse et al. 2008; Ku et al. 2008; Meissner et al. 2008; Mohn et al. 2008). It has been envisioned that, CpG island provides a specialized platform for coordinating transcription regulatory machinery. Molecular events occurring on this platform directly influence the cellular epigenetic landscape. (Blackledge and Klose 2011; Deaton and Bird 2011).

Histone modifications

In addition to DNA, core histones are subjected to covalent modifications. Each core histone consists of a globular histone fold domain and a protruding unstructured N-terminal “tail” (Luger et al. 1997; Kornberg and Lorch 1999). Post-translational modifications, such as acetylation, methylation, phosphorylation and ubiquitylation, occur predominantly in the tail region at specific residues. Most histone modifications have been shown to be reversible, and their physiological levels are balanced by enzymes establishing and removing the modifications. Histone modifications function either in mediating the establishment of specialized chromatin environments or coordinating with other factors in DNA-involved processes (Kouzarides 2007).

Currently, there are two proposed mechanisms for the functionality of histone modifications (Goldberg et al. 2007; Kouzarides 2007). First, modified histones can

influence the higher-order structure of chromatin (Tremethick 2007; Henikoff and Shilatifard 2011). Histone acetylation seems to facilitate transcription activation in this manner. The acetyl moiety neutralizes the positive charge of lysine residues, weakening DNA-histone interaction and thereby rendering chromatin an “open” structure (Shogren-Knaak et al. 2006). Second, histone modifications could recruit specific binding proteins that mediate downstream functions. Individual or combinational modifications have been hypothesized to serve as a functional “code” in establishing the specialized chromatin structure and mediating chromatin-related processes (Strahl and Allis 2000). For instance, it has been shown that H3K9me3, a modification enriched in heterochromatin, recruits HP1 protein via its chromodomain, which stabilizes the formation of heterochromatin (Lachner et al. 2001; Nakayama et al. 2001). H3K27me3, a repressive mark, is bound by Pc, a chromodomain-containing subunit of the PRC1 complex, which further carries out histone H2A ubiquitylation and reinforce repression (Cao et al. 2002; Fischle et al. 2003). H3K4me3, an active mark, specifically interacts with the PHD domain-containing protein in the NURF chromatin-remodeling complex, thereby tethering the remodeling activity to local chromatin and facilitating gene expression (Li et al. 2006; Wysocka et al. 2006).

Histone variants

The majority of nucleosomes are composed of canonical core histones, whose syntheses are tightly coupled to DNA replication. In contrast, non-canonical histone variants are constitutively expressed throughout cell cycle. These variants with unique amino acid compositions and modifications have specialized function in regulating chromatin structure and facilitating DNA-based processes (Banaszynski et al. 2010; Talbert and Henikoff 2010). For example, CENP-A, centromere-specific H3

preferentially binds to AT-rich DNA (as in centromere) with its specialized N-terminal tail, and its incorporation into nucleosome leads to atypical DNA bending, and affects nucleosome configuration. Such a specialized property is essential for the integrity of kinetochore (Black and Bassett 2008). H3.3 and H2A.Z, two variants with slight amino acid alterations from core histones, appear to have dual functions. Both H3.3 and H2A.Z associate with transcription activation in euchromatin; meanwhile, they localizes in pericentromeric regions and/or inactivated X chromosome (Marques et al. 2010; Szenker et al. 2011). MacroH2A, an H2A variant with a large C-terminal extension of non-histone domain, also plays a role gene suppression in inactivated X chromosome (Gamble and Kraus 2010). H2A.X, particularly in its phosphorylated form γ H2A.X, is specialized for signaling the recruitment of DNA repair machinery and nucleosome remodeling complexes upon double strand break (van Attikum and Gasser 2009). In addition, deposition of these variants to the chromatin requires specific histone chaperones. Interestingly, many histone variants and their deposition are dynamically regulated during embryogenesis. Depletions of these variants or chaperones lead embryonic lethality, suggesting essential roles of variant histones in development (Banaszynski et al. 2010).

Chromatin remodeling

Nucleosome positioning in the genome is dynamically regulated by ATP-dependent chromatin remodelers. Chromatin remodelers, which eject and reposition nucleosomes, are able to modulate local chromatin structure and control the accessibility of DNA to protein factors (Saha et al. 2006; Cairns 2009). These remodelers usually exist as large complexes. Based on their ATPase subunits, they can be divided into 4 families

based — SWI/SNF, ISWI, CHD and INO80. The assembly of chromatin remodeling complex can be diversified by selective incorporation of non-ATPase subunits (Wu et al. 2009; Ho and Crabtree 2010). For example, the BAF complexes of the SWI/SNF family have different assemblies along differentiation, due to the differential expression patterns of subunits. The incorporation of different subunits leads to the specialized functions in different differentiation stages (Lessard et al. 2007; Yan et al. 2008; Ho et al. 2009). Therefore, precise coordination of chromatin remodeling activities contributes to the cellular epigenetic status.

Noncoding RNA

In eukaryotic cells, most of the genome is transcribed, resulting in a large number of noncoding RNAs (ncRNAs). Some of these molecules have been shown to modulate chromatin dynamic and function in epigenetic regulation (Mercer et al. 2009; Wilusz et al. 2009). A well-known example is Xist and Tsix, two noncoding transcripts involved in X chromosome inactivation. Xist initiates silencing by binding to the X chromosome inactivation center (Xic) on one of the X chromosomes. Silencing is further reinforced by repressive histone modifications, such as H3K9me3 and H3K27me3. On the other hand, Tsix is responsible for keeping the active X chromosome from silencing. Tsix can titrate the histone modifying enzymes, or deplete Xist by annealing and triggering its degradation via RNA interference (RNAi) pathway (Heard and Disteche 2006; Yang and Kuroda 2007; Mercer et al. 2009). Furthermore, ncRNAs have also been shown to function in enforcing gene silencing in heterochromatin by coupling to RNA processing pathways. (Bernstein and Allis 2005; Zaratiegui et al. 2007). In addition, ncRNAs directly regulates transcription by other mechanisms: transcription from upstream ncRNA

promoter can interfere with the assembly of transcription machinery at the downstream promoters; ncRNAs transcribed by proximal promoters can recruit RNA-binding proteins that regulate transcription activity of the target promoter; and ncRNAs can serve as a cofactor stimulating transcription of adjacent promoter (Mercer et al. 2009; Wilusz et al. 2009).

To summarize, epigenetic regulations can be carried out with a variety of molecular mechanisms. How these mechanisms are involved in cell type specification and cell fate transition requires case-specific investigation.

1.3 Genetic and epigenetic regulation of embryonic stem cells

Mammalian development begins when two gametes, sperm and egg, are fused into a zygote. Zygote subsequently divides by cleavage into blastomeres. Both the zygote and early blastomeres are deemed as totipotent cells for their capacity to generate the whole organism; however, they lack the ability to self-renew. As development proceeds, cleavage division ends and cells acquire normal cell cycle properties in the blastocyst stage. Concurrent with that, first lineage specification takes place, as an outer layer of trophoblast and an inner cell mass (ICM) are formed. Trophoblast develops into extraembryonic tissues, including placenta, while ICM gives rise to the embryo proper. Cells from ICM are deemed as pluripotent cells for their capacity to differentiate into all adult cell types. These cells can acquire self-renewal property in specific culture conditions in vitro. The derived embryonic stem (ES) cells can be propagated indefinitely while maintaining their differentiation potentials. Molecular studies on ES cells show that

their unique property is controlled by a specific transcription factor circuitry that intertwines with epigenetic regulators.

Transcription factor circuitry for pluripotency

Transcription factors activate or repress gene transcription by recognizing their target DNA sequences. Oct4, Sox2 and Nanog are considered the core transcription factors governing the pluripotent state of ES cells (Jaenisch and Young 2008; Young 2011). Oct4 and Nanog are specifically enriched in ES cells compared to somatic cells (Nichols et al. 1998; Chambers et al. 2003; Mitsui et al. 2003). Oct4 is crucial for establishing pluripotency in ICM and ES cells (Nichols et al. 1998; Niwa et al. 2000), while Nanog is required for robust maintenance of pluripotency in ES cells and proper development after ICM is formed (Mitsui et al. 2003; Chambers et al. 2007; Silva et al. 2009). Sox2 heterodimerizes with Oct4 and functionally contributes to pluripotency (Ambrosetti et al. 2000; Avilion et al. 2003; Masui et al. 2007). The core transcription factor triad colocalize with each other to regulate a set of gene loci in ES cells (Boyer et al. 2005; Loh et al. 2006; Chen et al. 2008). First, these loci include the promoters of these core factors. The self-regulatory positive feedback loop sustains the robust expression of these core factors in ES cells. Second, Oct4, Sox2 and Nanog cooperate to activate genes expressed in ES cells. More than 60% of active genes in ES cells are target of these factors (Young 2011). Third, paradoxically, these core factors also localize to a set of lineage-specific genes, whose expression is repressed or “poised” in ES cells but rapidly activated upon differentiation. On these loci, these core factors cooperate with epigenetic factors, such as Polycomb complexes and SetDB1, which enforce the

repression (Boyer et al. 2006; Lee et al. 2006; Bilodeau et al. 2009). It has been thought that such a mechanism is crucial for maintaining the undifferentiated state of ES cells.

A series of transcription factors have been found to collaborate with the core factors, constituting an ES cell-specific transcription circuitry. Tcf3, Stat3, Smad1, which are components of Wnt, LIF and BMP pathway respectively, are implicated to coregulate gene expression with the core factors in ES cells. These factors incorporate external signals from the external environment to the transcription circuitry (Chen et al. 2008; Cole et al. 2008; Tam et al. 2008). c-Myc, which regulates proliferation and whose binding correlates with activation, also colocalizes with the core transcription factors in ES cells, suggesting a role in shaping the transcription output (Kim et al. 2008). In addition, transcription factors Sall4, Esrrb, Zfx, Tbx3, Rex1 and Klf4 have also been connected to the circuitry (Chen et al. 2008; Kim et al. 2008). Such a transcription factor network further outreaches to epigenetic regulators, which play special roles in maintaining ES cell identity (Orkin and Hochedlinger 2011; Young 2011).

Epigenetic regulation of embryonic stem cells

In general, ES cells have “open” chromatin, which is more permissive for transcription. It has been observed that heterochromatin is progressively clustered and rearranged into foci during differentiation (Wiblin et al. 2005; Meshorer et al. 2006; Williams et al. 2006). Chromatin components and chromatin associated proteins, such as histones H2B and H3, linker histone H1 and heterochromatin protein HP1, have been shown to exchange more vigorously in ES cells than in differentiated cells, suggesting an open state for ES cell chromatin (Meshorer et al. 2006). Such a chromatin state, as well

as the unique ES cell property, results from the specialized epigenetic regulatory mechanisms in ES cells.

A variety of epigenetic factors, including those mediating histone modifications, DNA methylation and chromatin remodeling, have been found to participate in epigenetic regulation in ES cells (Jaenisch and Young 2008; Meissner 2010; Orkin and Hochedlinger 2011; Young 2011). One of the mechanisms to maintain the ES cell property is through the establishment of “bivalent” chromatin structure (Boyer et al. 2005; Azuara et al. 2006; Bernstein et al. 2006; Lee et al. 2006). The bivalent structure is defined by the coexistence of the repressive mark H3K27me3 and the active mark H3K4me3. In ES cells, the bivalent structure covers a large set of differentiation-induced genes, whose expression are silenced or “poised” in ES cells. Upon differentiation, the bivalency on these loci is resolved and these genes are activated. In ES cells, H3K27me3 and/or its catalyzing polycomb repressive complex 2 (PRC2) dominates the activating mechanisms and keeps the expression of the differentiation-inducing genes poised. Although the activity of PRC2 is dispensable for pluripotency maintenance in ES cells, it is required for proper differentiation of ES cell (Chamberlain et al. 2008). Consistent with that, PRC2 components are crucial for development, as depleting any of them leads to early embryonic lethality (Shumacher et al. 1996; O'Carroll et al. 2001; Pasini et al. 2004).

In addition to PRC2, other histone modifying enzymes catalyzing repressive modifications are also implicated in functioning in ES cell maintenance. polycomb repressive complex 1 (PRC1), which catalyzes monoubiquitylation of H2A at lysine 119, shares a significant portion of gene targets with PRC2 in ES cells, including the lineage-

inducing genes (Boyer et al. 2006; Bracken et al. 2006). Setdb1, a histone methyltransferase specific for H3K9me3, has been shown to repress genes specific for extraembryonic trophoblast lineage in ES cells (Bilodeau et al. 2009; Yuan et al. 2009). Furthermore, the trithorax (Trx) complexes, which catalyze H3K4me3, play an important role in ES cell differentiation and interact with the core transcription factors (Ang et al. 2011; Jiang et al. 2011). The Tip40-p300 complex harboring histone acetylation activity has been found to be required for ES cell maintenance (Fazzio et al. 2008). Thus, histone modifying enzymes carry out diverse functions in regulating ES cell identity.

In ES cells, DNA methylation pattern shows correlations with histone modifications. Promoters with high CpG content, including those of pluripotency genes, lack DNA methylation and bear H3K4me3; while promoters with low CpG content, including those of lineage specific genes, are hypermethylated and devoid of H3K4me3 (Mikkelsen et al. 2007; Meissner et al. 2008). Such a pattern indicates that DNA methylation likely contribute to the ES cell specific transcription program that controls the ES cell state. Although ES cells deficient in DNA methylation can be derived and propagated, these cells cannot differentiate properly (Jackson-Grusby et al. 2001; Jackson et al. 2004), suggesting a role of DNA methylation in directing differentiation.

Chromatin remodeling complexes also play a role in regulating ES cell property. An ES-cell specific SWI/SNF family remodeling complex, esBAF, interacts and colocalizes with the core transcription factors. The components of esBAF are important for ES cell maintenance (Ho and Crabtree 2010). Chd1, a CHD family remodeler, localizes onto euchromatin and contributes to transcription activation in ES cells. Depletion of Chd1 in ES cells results in preferential differentiation to the neural lineage

(Gaspar-Maia et al. 2009). Chd7, another CHD remodeling enzyme, binds to genomic loci coregulated by the core transcription factors (Schnetz et al. 2010). All these evidences suggest that epigenetic enzymes modulating chromatin play important roles in specifying the ES cell state.

1.4 Generation of induced pluripotent stem (iPS) cells

Somatic cell reprogramming to pluripotency

It had been long thought that differentiated somatic cells achieve static cell fates by loss of chromosome or permanent gene inactivation during development, before somatic cell reprogramming to an embryonic state was achieved. Currently, there are three approaches for somatic cell reprogramming: nuclear transfer, cell fusion and transcription factor-directed reprogramming (Stadtfield and Hochedlinger 2010; Yamanaka and Blau 2010). In nucleus transfer, the nucleus from a somatic cell is transplanted into an enucleated oocyte, and the nucleus reprogramming is initiated by the oocyte-derived factors in the cytoplasm. The nuclear transferred oocyte is capable of developing into an entire individual, which is a genetically identical clone to the original somatic cells (Wilmut et al. 1997; Wakayama et al. 1998). A second approach to reprogram somatic cells to pluripotency is fusing somatic cells with pluripotent cells. After fusion, the pluripotent cell fate appears to dominate the differentiated one, as the nucleus from somatic cells acquires pluripotency in the resulting heterokaryon or hybrid cells (Tada et al. 1997; Tada et al. 2001). Recently, a third approach has been established. Pluripotency can be induced from somatic cells by ectopic expression of a set of transcription factors (Takahashi and Yamanaka 2006). The resultant induced pluripotent

stem (iPS) cells are nearly identical to ES cells derived from the ICM of blastocysts (Hanna et al. 2010; Stadtfeld and Hochedlinger 2010; Yamanaka and Blau 2010).

Transcription factor-directed iPS cell generation

Starting from 24 ES cell-enriched candidate factors, Yamanaka and colleagues found that overexpression of Oct4, Sox2, Klf4 and c-Myc is able to convert differentiated fibroblasts to an ES cell-like state. The derived iPS cells have ES cell molecular profiles and contribute to tissues in developing embryo after injected to blastocysts (Takahashi and Yamanaka 2006). Subsequently, iPS cells have been shown to be capable of generating chimera and transmitting to germ lines, when core transcription factor, Nanog or Oct4, is used as a reporter for pluripotency (Wernig 2004; Maherali et al. 2007b; Okita et al. 2007). Later, “all iPS-cell” mice have been generated through tetraploid complementation, which is the most stringent test for pluripotency (Boland et al. 2009; Kang et al. 2009; Zhao et al. 2009), illustrating that iPS cells derived from transcription factor-directed reprogramming bear developmental potential equivalent to ES cells.

This reprogramming strategy has been successfully applied to a wide range of somatic cell types, including terminally differentiated B cells, and across several mammalian species, including human (Stadtfeld and Hochedlinger 2010). A range of technical approaches have also been developed to introduce the transcription factors, including transduction by integrating retrovirus and lentivirus, transduction by non-integrating adenovirus, plasmid transfection, RNA transfection and protein delivery (Stadtfeld and Hochedlinger 2010).

Mechanisms of iPS cell generation

Models for the iPS cell generation process

Although the derivation of iPS cells is technically well-established, the mechanisms underlying the process have just begun to reveal. One of the most puzzling issues in iPS cell generation is its low efficiency. The conversion rate from starting cells to iPS cells is usually less than 1%, and the reprogramming process takes at least 1 to 2 weeks. To explain this inefficiency, two models have been raised: the “stochastic” model and the “elite” model (Hanna et al. 2009; Yamanaka 2009; Hanna et al. 2010; Stadtfeld and Hochedlinger 2010).

The stochastic model proposes that under the induction of transcription factors all somatic cells have equal potential to become iPS cells, and the reprogramming process must go through some stochastic events. The strongest evidence for this model comes from an elegant study starting with “secondary” lineage purified B cells, which inherently harbors inducible reprogramming factors. By monitoring for 18 to 20 weeks, the authors showed that nearly all the starting cells are able to generate a population of iPS cells, suggesting that nearly all starting cells have the potential to acquire pluripotency (Hanna et al. 2009). By mathematical modeling, the author further revealed a sole energy barrier during the reprogramming process (Hanna et al. 2009; Hanna et al. 2010). However, despite this evidence, a pure stochastic model seems difficult to explain some reprogramming observations which could be readily interpreted with the elite model.

The elite model proposes that certain cells in the reprogramming cell populations have advantages for becoming iPS cells. The advantages can be originated from the

privileged statuses of some cells among the starting cells. These statuses may include differentiation potentials, senescence status and etc. Evidences show that, reprogramming starting with hematopoietic stem or progenitor cells has significantly higher efficiency than those with fully differentiated hematopoietic cells (Eminli et al. 2009), and senescent cells display less reprogramming efficiency cells compared to the proliferative counterparts (Utikal et al. 2009). Furthermore, the advantages may also be gained by certain cells during the reprogramming process. It has been shown that at the beginning of reprogramming some population of cells undergoes rapid changes of morphology and proliferation profile, which leads to a high conversion rate to iPS cells compared to the overall population (Smith et al. 2010). For these observations, the elite model provides a simple and direct explanation. To reconcile all the reprogramming observations, it has been suggested that a model with elements from both the stochastic and elite model maybe more accurately reflects the iPS cell generation process (Stadtfield and Hochedlinger 2010).

Generation of iPS cells is a stepwise process

Generation of iPS cell has been portrayed as a stepwise process with multiple “roadblocks”, and somatic cells driven by the transcription factors must conquer these roadblocks to achieve pluripotency. It has been thought that, along the reprogramming process, fewer and fewer cells manage to pass each roadblock, which leads to the inefficiency of the final yield of iPS cells (Stadtfield and Hochedlinger 2010; Plath and Lowry 2011). Induction of proliferation, upregulation of epithelial genes and downregulation of some somatic cell gene are among the early events occurring during reprogramming. By time-lapse imaging, cells that successfully become iPS cells were

traced back to a population of cells that acquire a high proliferation rate and a compact size within the first day in the reprogramming process (Smith et al. 2010). Epithelial genes, which is expressed in ES cells but silenced in fibroblasts, have been shown to be activated early in reprogramming, which leads to a radical change in adhesion property and morphology of the reprogramming cells (Li et al. 2010; Samavarchi-Tehrani et al. 2010). By cell sorting, downregulation of fibroblasts surface marker Thy1 has also been shown to occur in the early stage of reprogramming (Stadtfeld et al. 2008). Cells that pass all these early roadblocks gradually acquire pluripotency, sequentially activating ES cell markers alkaline phosphatase, SSEA-1, Nanog and endogenous Oct4 (Brambrink et al. 2008; Stadtfeld et al. 2008). Once the transcription circuitry for pluripotency is established, the exogenous transcription factors become dispensable. Further “maturation” steps are probably needed for the nascent iPS cells to reactive inactivated X chromosome (Stadtfeld et al. 2008), erase the epigenetic memory of somatic cells completely (Polo et al. 2010) and fine-tune the pluripotent circuitry (Samavarchi-Tehrani et al. 2010).

Roles of the introduced transcription factors

Revealing the roles of the introduced transcription factors in iPS cell generation is essential for illuminating the reprogramming mechanisms. By genome wide localization studies, it has been found that the introduced factors bind to targets similar to those in ES cells (Sridharan et al. 2009a). Oct4 and Sox2, members of the core transcription triad in ES cells, bind to ES-cell enriched gene in reprogramming, and Klf4 collaborate with them by co-occupying half of their targets (Jiang et al. 2008; Kim et al. 2008; Sridharan et al. 2009a). Collaboration of these three factors is thought to be crucial for establishing the pluripotency network. Furthermore, Klf4 individually can activate some of the

epithelial genes, whose activation occurs early in reprogramming, suggesting a specialized role of in acquiring the epithelial property (Li et al. 2010).

Unlike Oct4, Sox2 and Klf4, c-Myc appears to play a different role in reprogramming. It binds to a distinct set of targets that function in cell proliferation, metabolism and biosynthetic pathway and functions early in reprogramming (Mikkelsen et al. 2008; Sridharan et al. 2009a). Furthermore, although it greatly enhances iPS cell generation, c-Myc is dispensable for iPS cell derivation (Nakagawa et al. 2008; Wernig et al. 2008). c-Myc has been shown to promotes transcription elongation (Rahl et al. 2010) and facilitate the establishment of active chromatin environment (Knoepfler 2008). Therefore, it is likely that c-Myc generally contributes to gene activation, which is directed by other introduced reprogramming factors.

Additional factors that affect iPS cell generation

Currently, a variety of additional factors, including transcription factors, chromatin modulators, miRNAs, growth factors and chemical compounds have been found to affect the reprogramming process (Feng et al. 2009; Stadtfeld and Hochedlinger 2010). Understanding how these factors contribute to reprogramming helps uncover the mechanisms of iPS cell generation.

Nanog, despite its essential role in maintaining pluripotency in ES cell, was not among the factors originally identified by Yamanaka and colleague (Takahashi and Yamanaka 2006). However, later, it has been found that, although Nanog is not required for the initiation of reprogramming, it facilitates the transitions from intermediate cells to iPS cells and enhances reprogramming efficiency (Hanna et al. 2009; Silva et al. 2009).

In addition, Nanog has been identified in the cocktail of transcription factors inducing human iPS cells, together with Oct4, Sox2 and Lin28 (Yu et al. 2007), again suggesting an important role of Nanog in establishing pluripotency in iPS cell.

Modulations of cell proliferation status have also been found to affect reprogramming. Depletion of p53-p21 pathway and suppression of Ink4a/Arf accelerate the reprogramming kinetics and enhance reprogramming efficiency (Banito et al. 2009; Hanna et al. 2009; Hong et al. 2009; Kawamura et al. 2009; Li et al. 2009; Utikal et al. 2009). Enhanced proliferation has been suggested to facilitate epigenetic changes presumably by efficient resetting the chromatin states in S phase (Hanna et al. 2009; Plath and Lowry 2011).

Given that acquiring epithelial properties and eliminating mesenchymal ones are among the steps in reprogramming from fibroblasts to iPS cells, factors promoting mesenchymal-epithelial transition (MET) and/or inhibiting epithelial-mesenchymal transition (EMT) are capable of facilitating reprogramming. These factors include agonists of BMP pathways, inhibitors of Tgf- β pathway and miRNA clusters miR-200s (Ichida et al. 2009; Maherali and Hochedlinger 2009; Li et al. 2010; Samavarchi-Tehrani et al. 2010).

Factors that target chromatin have been also shown to play a role in iPS cell generation. Components of the BAF chromatin remodeling complex are capable of enhancing reprogramming efficiency (Singhal et al. 2010). Chemicals that inhibit DNA methylation, histone deacetylation or histone H3 K9 methylation are also capable of promoting iPS cell generation (Huangfu et al. 2008a; Mikkelsen et al. 2008; Shi et al.

2008a). Interestingly, all these epigenetic modulators leads to a permissive chromatin, suggesting that counteracting the suppressive chromatin state in somatic cell may be a crucial step for reprogramming.

To summarize, various additional factors acts on the reprogramming process from different aspects. Identifying novel factors that contribute to iPS cell generation is one of the strategies to understand the mechanisms of reprogramming.

1.5 Introduction to the research within the dissertation

Generation of iPS cells attracts huge attention for its tremendous clinical potentials in regenerative medicine; however, the low efficiency and the insufficient mechanistic knowledge of the reprogramming process cast uncertainty on the future application of iPS cells. To improve the iPS cell derivation process and advance our understanding of reprogramming mechanisms, I aimed to identify epigenetic modulators that facilitate iPS cell generation, based on the premise that all cell fate transitions, including reprogramming to iPS cells, are fundamentally epigenetic processes.

I first focus on the chemical epigenetic modulators and found that butyrate, a histone deacetylase inhibitor, promotes iPS cell generation, Butyrate not only increases iPS cell number and changes the reprogramming dynamics, but also reduces the frequency of partially reprogramming cells. The facilitation by butyrate occurs early in reprogramming and depends on the presence of c-Myc. Genome-wide expression study reveal the upregulation of ES cell enriched genes in the presence of butyrate. This research strengthens the viewpoint that reversing the repressive chromatin structure is

crucial for reprogramming to iPS cells. This part of my research will be presented in Chapter Two.

Next I probed for the potential effect of ES cell-enriched epigenetic factors, which potentially help establish the ES cell state, and found that Kdm2b, a histone demethylase specific for H3K36me₂, is capable of enhancing iPS cell generation. Details studies reveal that, the Kdm2b-directed enhancement is independent of its role in promoting cell proliferation or suppressing senescence, but relies on its effect on activating early epithelial transcription program. Microarray and gene ontology analysis indicates that there likely exists a switch from an epithelial transcription program to a pluripotent program during iPS cell generation. For that, I propose a transcription cascade model for iPS cell generation and address the role of Kdm2b using this model. This part of my research will be elaborated in Chapter Three.

Overall, research in this dissertation not only shows that manipulation of epigenetic statuses is a way to improving iPS cell generation, but also shed light on the reprogramming mechanisms by highlighting the importance of epigenetic modifications.

CHAPTER 2 BUTYRATE PROMOTES INDUCED PLURIPOTENT STEM CELL GENERATION*

2.1 Introduction

Reprogramming from somatic cells to induced pluripotent stem (iPS) cells can be achieved by retroviral expression of transcription factors Oct4, Sox2, Klf4 and c-Myc (Takahashi and Yamanaka 2006; Maherali et al. 2007a; Okita et al. 2007; Wernig et al. 2007). However, the slow reprogramming process and low reprogramming efficiency impede detailed mechanistic studies and potential applications of this technology. One solution to overcome these problems is to identify small molecules that can enhance the reprogramming efficiency. Indeed, recent studies have demonstrated that several chemicals with the capacity to modulate epigenetic enzymes exhibit positive effects on iPS cell generation. For example, valproic acid (VPA), a histone deacetylase inhibitor, has been shown to improve both the kinetics and efficiency of mouse and human iPS cell generation (Huangfu et al. 2008a; Huangfu et al. 2008b). In addition, BIX-01294, an inhibitor for the histone methyltransferase G9a, and RG108, a DNA methyltransferase (DNMT) inhibitor, have been reported to enhance the efficiency of iPS cell generation (Shi et al. 2008a; Shi et al. 2008b). Furthermore, another DNMT inhibitor, 5-aza-cytidine, has been shown to facilitate the conversion of partially reprogrammed cells to fully reprogrammed iPS cells (Mikkelsen et al. 2007).

* Adapted from Liang G, Taranova O, Xia K, Zhang Y. 2010. *J Biol Chem* **285**: 25516-25521.

Recently, butyrate, a naturally occurring short-chain fatty acid and histone deacetylase inhibitor, has been shown to support self-renewal of both human and mouse embryonic stem (ES) cells in a range of relatively low concentrations (Ware et al. 2009). However, butyrate has also been reported to induce differentiation when applied at higher concentrations (Newmark et al. 1994). Therefore, whether butyrate has an effect on iPS cell generation is an intriguing question. In this report, we sought to examine the effect of butyrate on iPS cell generation. We found that butyrate facilitates iPS cell generation in the range of 0.5 to 1 mM. This effect appears to be mediated through one of the reprogramming factors c-Myc. In addition, butyrate is able to increase the percentage of fully reprogrammed iPS cells by reducing partially and/or unsuccessfully reprogrammed cells. Genome-wide gene expression analysis indicates that butyrate can specifically increase the expression of some ES cell-enriched genes in fibroblasts in the presence of exogenous c-Myc. Thus our studies uncover another chemical capable of facilitating iPS cell generation, contributing to the iPS cell tool box.

2.2 Results

Butyrate promotes iPS cell generation

A recent study indicated that butyrate, a small fatty acid, supports self-renewal in mouse and human embryonic stem cells (Ware et al. 2009). To determine whether butyrate has an effect in iPS cell generation, we transduced 1×10^5 MEFs derived from hemizygote Sox2-GFP mice with retroviruses expressing Oct4, Sox2, Klf4 and c-Myc in the presence of varying concentrations of butyrate. The effect of butyrate on reprogramming was monitored for a period of 12 days after infection. In the presence of

butyrate, GFP⁺ colonies with ES cell-like morphology were observed at day 6 post-infection (Figure 2-1A). At day 8, we observed a dose-dependent enhancement of reprogramming efficiency when butyrate was used at concentrations between 0.25–1 mM. A maximum of 7-fold increase was observed when butyrate was applied at 1 mM concentration. However, at higher concentrations (1.5 and 2 mM), butyrate becomes cell toxic and no GFP positive colonies were observed (Figure 2-1B). On day 12, treatment of butyrate at 0.5 and 1 mM still showed an approximate 2-fold increase in the number of GFP positive colonies (Figure 2-1B, right panel, green bars). The effect of butyrate on promoting the generation of GFP positive iPS colonies is comparable to that of VPA (Figure 2-1B).

In addition to counting the GFP positive iPS cell colony numbers, we also counted the total and the alkaline phosphatase (AP) positive colony numbers on day 12. Interestingly, at lower concentrations of butyrate (0.25–1 mM), both the AP⁺ and total colony numbers show a concentration-dependent reduction (Figure 2-1B, right panel), although more GFP⁺ colonies were observed. This leads to exceptionally high GFP⁺/AP⁺ and GFP⁺/total colony ratios under butyrate treatment, which is distinct from the effect of VPA (Figure 2-1B, right panel). For example, in the absence of butyrate, only 20% of the total colonies are GFP⁺ and 63% are AP⁺; at the concentration of 0.5 mM, 66% are GFP⁺ and 83% are AP⁺; at the concentration of 1 mM, these ratios are further increased to 92% and 96% respectively (Figure 2-1B, right panel and Figure 2-1C). Since Sox2-GFP is a more stringent pluripotency marker than AP, the reduction of GFP⁻ colonies in total population and AP⁺ population suggests that butyrate is capable of suppressing the formation of transformed cells or partially reprogrammed cells that

were not destined to the pluripotent cell fate. This observation is consistent with the well-characterized effect of butyrate on limiting the cell growth of cancerous cells (Bolden et al. 2006).

Butyrate accelerates iPS cell generation and its effect is c-Myc dependent

To further characterize the effect of butyrate on iPS cell generation, we monitored the effect of butyrate (1 mM) on the kinetics of reprogramming by introducing the four transcription factors (4F; Oct4, Sox2, Klf4 and c-Myc) and 3 factors (3F; Oct4, Sox2 and Klf4). In the case of 4F reprogramming, butyrate accelerated formation of GFP⁺ colonies by 2-3 days until post-infection day 16 when the effect of butyrate on the GFP⁺ colony number becomes unnoticeable (Figure 2-2A). However, when the reprogramming was performed using three factors, butyrate appears to have a negative effect on reprogramming efficiency (Figure 2-2A). This suggests that the enhancement effect of butyrate on reprogramming is dependent on exogenous c-Myc under our experimental condition. We also monitored the number of AP⁺ colony and total colony at post-infection day 12. In both 4F and 3F reprogramming, butyrate strongly reduces the total colony number and GFP⁻ colony number (Figure 2-2B), indicating that butyrate is capable of limiting the formation of partially reprogrammed cells or transformed cells regardless of whether three factors or four factors were used for reprogramming. In addition, we also tested the potential effect of butyrate on reprogramming by withdrawing other factors from the 4F combination; however, we did not observe any positive effect of butyrate on reprogramming under these conditions (data not shown).

Sox2-GFP positive colonies are pluripotent

Next we set out to characterize the GFP⁺ colonies generated in the presence of butyrate. To this end, individual GFP⁺ colonies were picked-up and propagated in standard ES cell culture medium in the absence of butyrate. All of the GFP⁺ colonies derived in the presence of butyrate exhibit ES-like morphology (Figure 2-3A) and high alkaline phosphatase activity (Figure 2-3B). We randomly picked three colonies for further characterization. RT-PCR analysis demonstrated that all three lines expressed endogenous stem cell factors Oct4, Sox2, Nanog, as well as other pluripotency-related genes, such as Fbx15 and Utf1 (Figure 2-3C). When these iPS cells were injected into immunodeficient mice, all lines were able to form complex-structured teratoma containing tissues of the three germ layers (Figure 2-3D). Collectively, these results suggest that Sox2-GFP⁺ cells generated in the presence of butyrate are pluripotent.

Butyrate facilitates iPS cell generation in an early time window

To shed light on the role of butyrate in promoting iPS cell generation, we sought to determine the time window during which butyrate exerts its effect. First, butyrate was applied to the culture media immediately after infection and was withdrawn from the media at different time points during the 12-day reprogramming process (Figure 2-4A, left panel). Successfully reprogrammed GFP⁺ colonies were counted at day 12 and the effect of butyrate on reprogramming was determined by comparison to GFP⁺ colony numbers in the absence of butyrate. Results shown in Figure 2-4A (right panel) demonstrate that the exposure to butyrate for only 2–4 days following transduction has a similar effect as that of continued exposure during the reprogramming process. Next, we

determined whether the 2–4 days of butyrate treatment needed to be performed at a particular time point during the reprogramming process. To this end, butyrate was added to the culture media at different time points during the reprogramming process (Figure 2-4B, left panel) and the effect of butyrate on reprogramming was determined in a similar way as that described above. Interestingly, we found that exposure to butyrate for 2 days after transduction has the maximum positive effect on the efficiency of reprogramming. Based on the above experiments, we conclude that butyrate exhibits the maximum effect on iPS cell generation at the initial 2–4 days of reprogramming, suggesting that butyrate functions early during the reprogramming process. Our finding that butyrate exerts its effect in a c-Myc-dependent manner (Figure 2-2) and that this occurs early during the reprogramming process is consistent with previous studies demonstrating that c-Myc mainly contributes to reprogramming at an early stage (Sridharan et al. 2009b).

Butyrate up-regulates a set of ES cell-enriched genes in c-Myc-mediated reprogramming

To gain insight into the molecular mechanism of butyrate enhanced reprogramming, we carried out gene expression studies using cDNA microarrays with the following four samples: i) 4F transduction, ii) 4F transduction with butyrate treatment (4F+B), iii) 3F transduction, and iv) 3F transduction with butyrate treatment (3F+B). Because the effective time window for butyrate is 2-4 days after infection, and since we are mainly interested in a primary effect, we treated the transduced MEFs with butyrate for a period of 48 hrs prior to harvesting RNA. We first focused on the genes whose expression is at least 8-fold higher in ES cells than in MEFs (Mikkelsen et al. 2007). We plotted the expression level of these ES cell-enriched genes in 4F (Figure 2-5A, red dots)

and 3F (blue dots) transduced MEFs at day 4 on a scatter chart, in which x and y axis respectively represent the expression level in the absence or presence of butyrate (Figure 2-5A). Linear regression of the scatter plot show that the slope of the 4F regression line (1.3137) is significantly higher than that of the 3F regression line (1.0918), indicating that the expression of these stem cell enriched genes as a whole is increased by butyrate treatment in a c-Myc-dependent manner (Figure 2-5A). Similarly, we have also analyzed the expression of MEF-enriched genes in response to butyrate treatment in reprogramming. However, a significant down-regulation of these genes due to the treatment of butyrate was not noticed, regardless of whether 4F or 3F were used in reprogramming (data not shown).

Further analysis of the microarray data indicate that a total of 337 probes were at least 2-fold up-regulated by the treatment of butyrate in the 4F reprogramming, while only 182 probes were up-regulated at least 2-fold in 3F reprogramming (Figure 2-5B). Interestingly, 199 out of the 337 probes up-regulated in the 4F reprogramming appear to be c-Myc-dependent as butyrate treatment failed to significantly up-regulate them in the 3F reprogramming. Among the 199 probes, 21 probes correspond to 19 known ES cell-enriched genes (Table S2-1). RT-qPCR analysis of randomly selected genes out of the 19 ES cell-enriched genes confirmed that their expression is significantly up-regulated by the treatment of butyrate in the 4F reprogramming, but not in the 3F reprogramming (Figure 2-5C). The selected genes, *Bex1*, *Mreg* and *Krt12* are most up-regulated by butyrate at the concentration of 1 mM (Figure S2-1), which is consistent with the observed maximum effect of butyrate on iPS cell generation at this concentration. How

up-regulation of these genes contributes to the iPS cell generation process remains to be determined.

2.3 Discussion

In this study, we demonstrate that butyrate, a small fatty acid and histone deacetylase inhibitor, promotes mouse iPS cell generation by Oct4, Sox2, Klf4 and c-Myc. Under the optimal concentration (0.5–1 mM), butyrate can enhance the generation of Sox2-GFP positive iPS cells by 7-fold when the four factors were used in reprogramming (Figure 2-1B). Butyrate facilitates iPS cell generation mainly by shifting the kinetic of 4F reprogramming 2-3 days forward (Figure 2-2A). A recent study indicated that ES cell self-renewal is facilitated by the presence of butyrate at a lower concentration (around 0.2 mM) (Ware et al. 2009). Given that butyrate can relax the chromatin structure by functioning as a histone deacetylase inhibitor, a higher (0.5-1 mM) optimal concentration in 4F reprogramming might indicate that establishment of pluripotency may need more accessible chromatin structure compared with that required for maintenance of ES cell status. Although butyrate is widely used as a differentiation reagent when applied at a higher concentration (≥ 1 mM) (Newmark et al. 1994), we did not notice differentiation of the fully reprogrammed Sox2-GFP⁺ colonies during the time of butyrate treatment. Neither do we observe any negative effect of butyrate on the quality of iPS cells derived in the presence of butyrate. The apparent conflicting roles that butyrate displays in differentiation and iPS cell generation suggest that this epigenetic modulator may generally facilitate cell fate changes by increasing the flexibility of chromatin. In this regard, it will be interesting to test whether butyrate can facilitate

transdifferentiation between different cell types, for example from fibroblasts to muscle cells (Davis et al. 1987) and from B cells to macrophage (Xie et al. 2004).

In addition to facilitating iPS cell generation, we also noticed that butyrate can reduce the GFP negative colony numbers, regardless of their AP activity status. This effect results in a significant increase in the ratio of GFP⁺ colonies, representing authentically reprogrammed iPS cells (Fig 1B and C). The AP⁻ GFP⁻ colonies are probably cells that failed to express all four reprogramming factors. These cells usually exhibit properties of transformed cells, such as granulated or cobblestone morphology with fast cell growth (Takahashi and Yamanaka 2006). Elimination of these AP⁻ GFP⁻ colonies can be attributed to the well-characterized anti-cancer effect of butyrate (Bolden et al. 2006), including induction of p21/Cdkn1a and p19/Arf (Figure S2-1). On the other hand, AP⁺ GFP⁻ colonies most likely represent partially reprogrammed cells that somehow have not achieved the pluripotent cell fate (Mikkelsen et al. 2007; Sridharan et al. 2009b). With the assistance of butyrate, these partially reprogrammed cells may gain full pluripotency and express Sox2-GFP contributing to the observed higher reprogramming efficiency. We note that although VPA, another histone deacetylase inhibitor, can also increase the GFP⁺ colony number, it does not seem to have the same capacity to suppress the total colony numbers (Figure 2-1B). Another epigenetic modulator, 5-aza-cytidine, has been shown capable of facilitating conversion of partially reprogrammed cells to fully reprogrammed cells (Mikkelsen et al. 2007). It will be interesting to test whether butyrate has a similar property. Given that butyrate functions at an early stage of reprogramming, while 5-aza-cytidine facilitates conversion of partially

reprogrammed cells to fully reprogrammed cells, it will be interesting to test whether they can function synergistically to facilitate the reprogramming process.

Since reprogramming is a process with multiple steps (Brambrink et al. 2008; Stadtfeld et al. 2008), we determined the functioning time window of butyrate to be at an early stage of the reprogramming process (Figure 2-4). This is consistent with our finding that butyrate facilitates reprogramming only in the presence of exogenous c-Myc (Figure 2-2) as c-Myc has been suggested to contribute to the early events of reprogramming (Sridharan et al. 2009b). To explore the effect of butyrate in the early reprogramming process at the molecular level, we analyzed the genome-wide expression profiles of cells undergoing reprogramming in this time window. We observed a trend of up-regulation for ES cell-enriched genes in response to butyrate (Figure 2-5A), but did not detect genome-wide down-regulation of MEF-enriched genes (data not shown). This observation is consistent with the role of butyrate in gene activation as a histone deacetylase inhibitor. Furthermore, we identified 19 ES-enriched genes that are specifically up-regulated by butyrate only when c-Myc is included as a reprogramming factor (Table S2-1). It remains to be determined whether up-regulation of these genes mediates the effect of butyrate. Given that reprogramming efficiency can be increased by suppression of the p53-p21 pathway (Hanna et al. 2009; Hong et al. 2009; Kawamura et al. 2009; Marion et al. 2009) as well as elimination of the senescence barrier imposed by Ink4a and/or Arf (Li et al. 2009; Utikal et al. 2009), we also analyzed whether butyrate can suppress the expression of p53, p21, Ink4a and Arf. Instead of down regulation, we indeed observed a slight up-regulation of some of these genes (p21 and Arf) by the

butyrate treatment (Figure S2-2), largely excluding the involvement of p53-p21 pathway in mediating the butyrate effects.

During the preparation of this manuscript, Mali et al. reported that butyrate greatly facilitates human iPS cell generation (Mali et al. 2010). In their report, reprogramming efficiency is increased remarkably by butyrate even in the absence of c-Myc and Klf4. Another difference between these two studies is the timing at which butyrate exhibits its effect. While it exerts an effect at an early stage during reprogramming in mouse cell reprogramming, Mali et al reported a later effect in human cells (Mali et al. 2010). Furthermore, while we noticed an inhibitory effect for transformed cells and partially reprogrammed cells, it is not clear whether a similar effect is seen in human cells. Whether these differences are due to the endogenous c-Myc levels, the different times required to achieve reprogramming for human and mouse MEFs, or other technical aspects remains to be determined. Nonetheless, the demonstration that butyrate, an HDAC inhibitor, is capable of facilitating iPS cell generation suggests that alteration of epigenetic status is an important step for the establishment of pluripotency.

2.4 Materials and methods

MEF derivation and iPS cell culture

Mouse embryonic fibroblasts (MEFs) were derived from E13.5 embryos of Sox2-GFP/Rosa26-M2rtTA double knock-in mice. MEFs were cultured in rich fibroblast growth medium (Dulbecco's modified Eagle's medium [DMEM] supplemented with 15% fetal bovine serum [FBS], non-essential amino acid [NEAA], GlutaMax, and penicillin/streptomycin) for no more than 2 passages before retroviral transduction. iPS

cells were cultured in mouse ES cell medium (DMEM with 15% FBS, NEAA, GlutaMax, sodium pyruvate, β -mercaptoethanol, penicillin/streptomycin and 1,000 U/ml leukemia inhibitory factors) with mytomycin C-treated STO cells as feeder cells or on 0.1% gelatin-coated plates.

Retrovirus preparation and infection

Retroviral plasmids pMXs expressing murine Oct4, Sox2, Klf4 and c-Myc (Takahashi and Yamanaka 2006) were transfected respectively into 293T cells with packaging plasmids pGag-pol and pVSVG. Virus-containing supernatants were harvested at 48 h and 72 h after transfection and were filtered through 0.45- μ m filter membrane and concentrated by spin column, before being used in MEF transduction in the presence of polybrene (5 μ g/ml).

Generation of iPS cells and calculation of reprogramming efficiency

MEFs were seeded in 6-well plates with 1×10^5 cells per well, 16 h before the first infection. Concentrated viruses were applied to MEFs twice within 48 h. The day when viral supernatant was removed was defined as day 0 post-infection. Transduced fibroblasts were then cultured in mouse ES cell medium in the presence or absence of butyrate in 12-16 days period. The concentration of butyrate used is 1 mM, except stated otherwise. In most cases, reprogramming efficiency is represented by the Sox2-GFP+ colony number derived from 1×10^5 MEFs; in some case, relative reprogramming efficiency is also used, which is the fold change of Sox2-GFP+ colony number with butyrate treatment compared to that without treatment. Alkaline phosphatase (AP) staining was performed with alkaline phosphatase detection kit (Millipore).

Teratoma formation and analysis

Teratomas were induced by subcutaneously injecting 1×10^6 iPS cells into Rag2^{-/-}; γ C^{-/-} immunodeficient mice. Xenografted tumor samples were isolated from mice in four to six weeks, fixed by 4% paraformaldehyde, embedded in paraffin and processed for hematoxylin and eosin staining, using standard protocols.

Quantitative and semi-quantitative RT-PCR

Total RNAs were harvested using RNeasy kit (Qiagen). Primers for quantitative and semi-quantitative RT-PCR are listed in Table S2-2. Quantitative PCR reactions were performed with SYBR GreenER mix (Invitrogen). Relative gene expression levels were normalized to *Gapdh* mRNA.

Genome-wide expression analysis

2 μ g of total RNA were reverse-transcribed into cDNA with a T7-(dT)₂₄ primer from a custom kit (Life Technologies). Biotinylated cRNA was then generated from the cDNA reaction using the BioArray High Yield RNA Transcript Kit. The cRNA was then fragmented in fragmentation buffer (40 mM Tris-acetate, pH8.1, 100 mM KOAc and 150 mM MgOAc) at 94 °C for 35 minutes before microarray hybridization. 15 μ g of fragmented cRNA was then added to a hybridization cocktail (0.05 μ g/ μ l fragmented cRNA, 50 pM control oligonucleotide B2, BioB, BioC, BioD, and cre hybridization controls, 0.1 mg/ml herring sperm DNA, 0.5 mg/ml acetylated BSA, 100 mM MES, 1 M [Na⁺], 20 mM EDTA, 0.01% Tween 20). 10 μ g of cRNA were used for hybridization to Affymetrix GeneChip Mouse Genome 430A 2.0 Array. Hybridization was carried out at 45 °C for 16 hours. The array was then washed and stained with R-phycoerythrin

streptavidin, before scanning. Washing, scanning and basic analysis was carried out using Affymetrix GeneChip Microarray Suite 5.0 software.

Western blotting

Cell populations were collected at day 2 post-infection and lysed with RIPA buffer (150 mM NaCl, 50 mM Tris, pH7.5, 1% NP40, 0.05% Na-deoxycholate, 1 mM Na₃VO₄ and 1 mM NaF) for 1 h at 4 °C. Soluble fraction, as total protein extract, was isolated by centrifugation and then subjected to electrophoresis. Western blotting was performed with antibodies against p53 (Santa Cruz, sc6243), p21 (Santa Cruz, sc52870), Ink4a (Santa Cruz, sc1207), Arf (Santa Cruz, sc32748) and α -tubulin (Sigma, T6199).

Figure 2-1 Butyrate promotes iPS cells generation.

(A) A representative Sox2-GFP positive iPS cell colony generated in the presence of butyrate at post-infection day 6. Scale bar, 100 μm . Mouse embryonic fibroblasts (MEFs) from hemizygote Sox2-GFP mouse were transduced with Oct4, Sox2, Klf4 and c-Myc (4F), and treated with butyrate. (B) GFP+ colony numbers counted at day 8 (left panel) post transduction of 1×10^5 MEFs with the four reprogramming factors in the absence or the presence of various concentration (0.25, 0.5, 1, 1.5 and 2 mM) of butyrate or VPA (1 mM). Presented in the right panel are the numbers of GFP-positive and total colony counted at day 12 post transduction. Subsequently, cells were stained for alkaline phosphatase (AP). The number of GFP-positive colonies (green), AP-positive GFP-negative colonies (purple), and AP-negative GFP-negative colonies (grey) are shown in the chart. (C) Representative fluorescent microscopic pictures taken at day 12 post transduction by 4F and 4F in the presence of 0.5 mM butyrate (4F+B). Arrows indicate Sox2-GFP positive colonies. Scale bar, 1 mm.

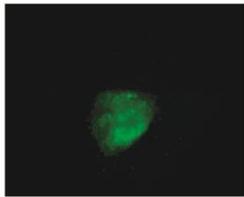
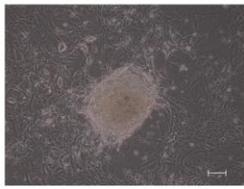
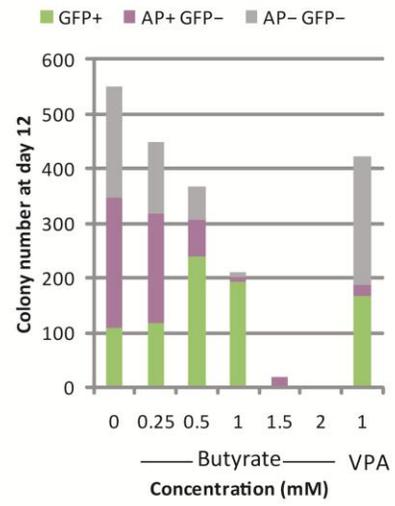
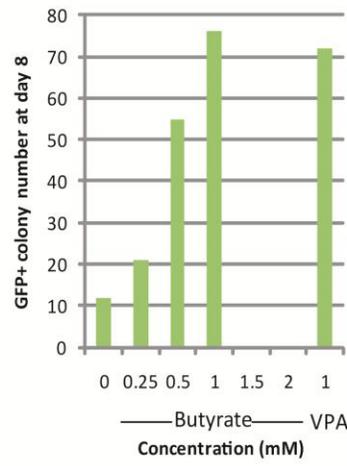
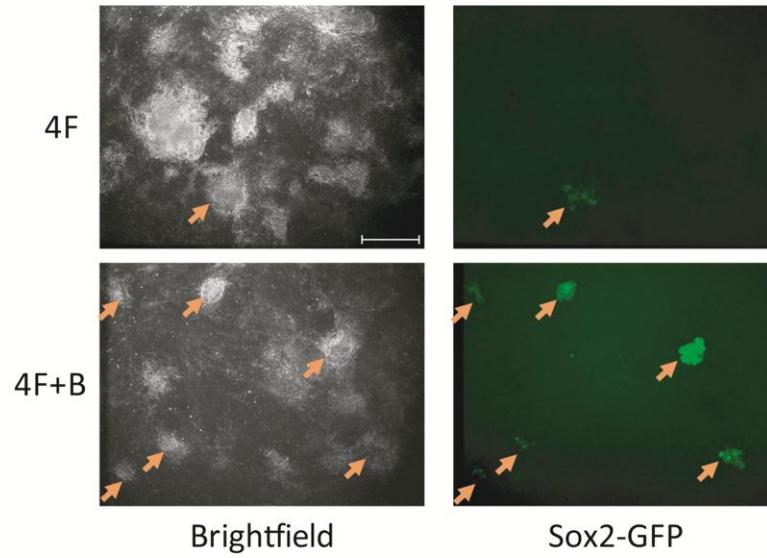
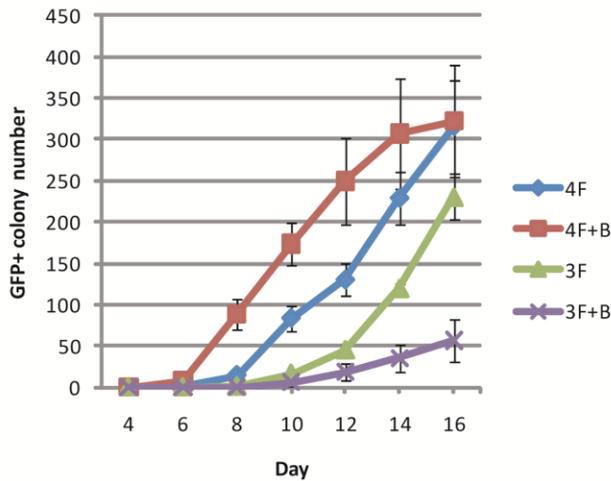
A**B****C**

Figure 2-2 Butyrate improves the reprogramming kinetics in a c-Myc-dependent manner.

(A) Kinetics of reprogramming in the presence of butyrate. Sox2-GFP MEFs transduced with 4F and 3F were treated with butyrate (1 mM). The number of GFP-positive colonies for each treatment was counted at different days until day 16 post-transduction. (B) GFP-positive colonies and total colonies were counted at day 12. AP-positive colonies were also counted, after cells were stained for AP.

A



B

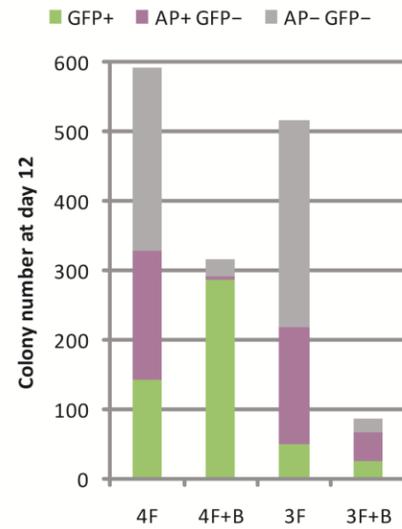


Figure 2-3 iPS cells generated in the presence of butyrate are pluripotent.

(A) Representative pictures of an iPS cell line derived in the presence of butyrate. iPS cell lines derived from 4F reprogramming in the presence of butyrate were cultured in ES cell medium on gelatin-coated plates without feeder cells. Scale bar, 100 μm . (B) iPS cells derived in the presence of butyrate is AP positive. Scale bar, 100 μm . (C) RT-PCR demonstrate that three randomly picked iPS cell lines derived in the presence of butyrate express pluripotent marker genes. (D) Representative pictures of teratoma derived from butyrate-assisted iPS cell lines comprise of cell types from all three germ layers (endoderm, mesoderm and ectoderm). Scale bar, 100 μm . All three randomly picked iPS cell lines showed a similar capacity in generating teratomas harboring cells belonging to the three germ layers.

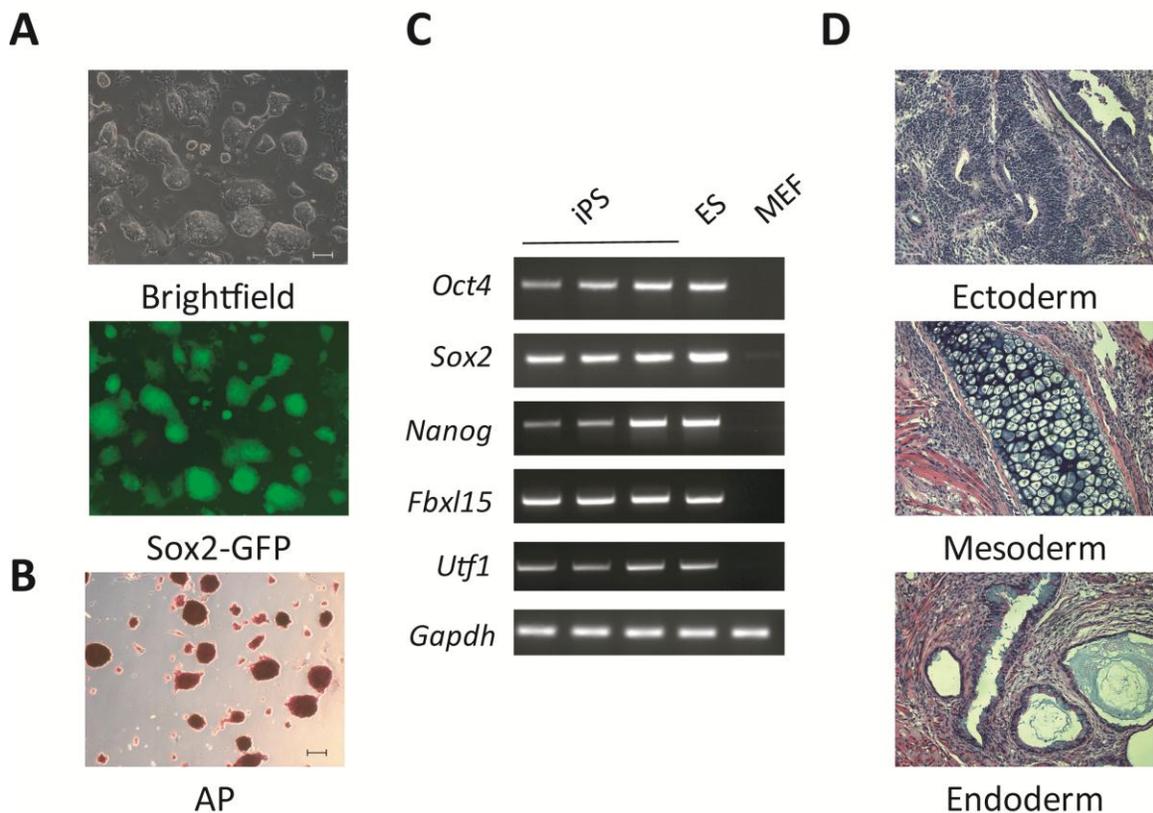


Figure 2-4 Butyrate facilitates iPS cell generation at an early time window during reprogramming.

(A) Sox2-GFP MEFs transduced with 4F were treated with butyrate immediately after transduction with 4F and butyrate was removed from the culture media at different time points and the GFP+ colonies are counted at day 12. The reprogramming efficiencies of various treatments were compared to that without the butyrate treatment and presented on the right panel. **(B)** Sox2-GFP MEFs transduced with 4F were treated with butyrate at different time after transduction with 4F and butyrate was maintained in the culture media till day 12 when the GFP+ colonies are counted. The reprogramming efficiencies of various treatments were compared to that without the butyrate treatment and presented on the right panel.

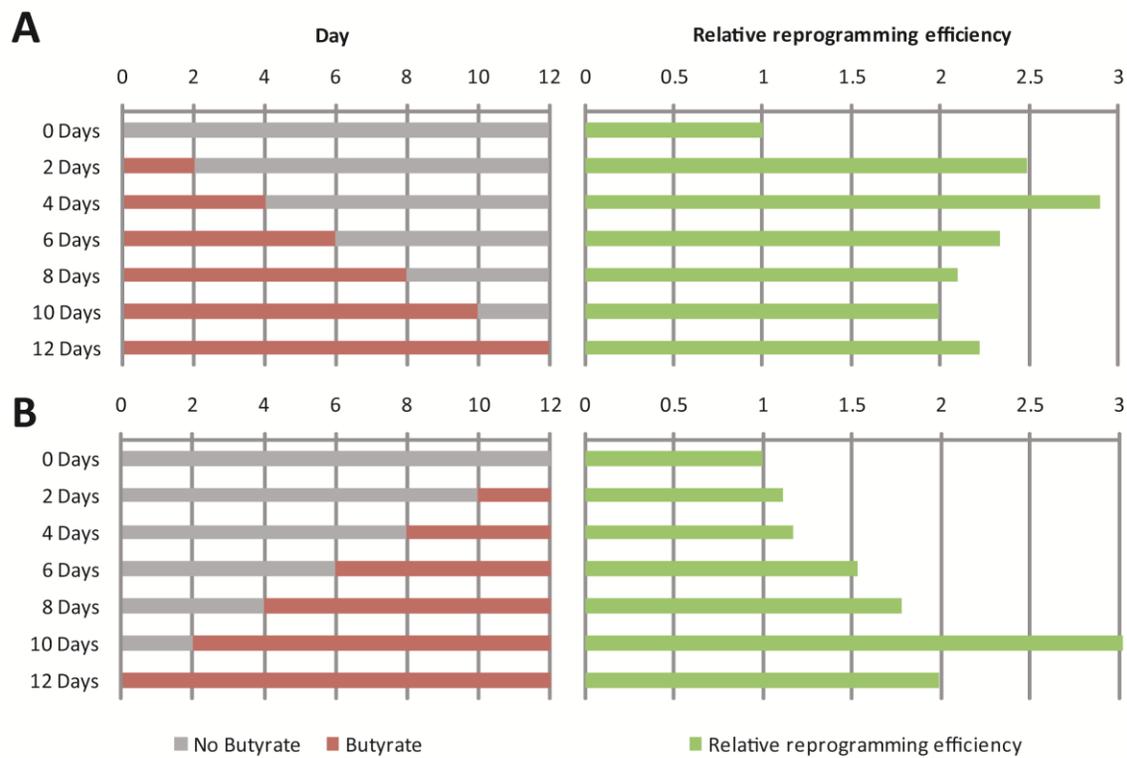


Figure 2-5 Butyrate enhances the expression of a set of ES cell-enriched genes in a c-Myc-dependent manner.

(A) Global gene expression was analyzed by microarray (Affymetrix) using total RNA samples harvested 2 days after mock or butyrate treatment. The expression levels of ES cell-enriched genes from samples with butyrate treatment (y-axis) are plotted against those without butyrate treatment (x-axis). The expression from cells transduced with 4F and 3F are respectively shown in red square and blue diamond. The linear regression line for 4F (red, $y = 1.3137x - 10.675$, $R^2 = 0.8705$) is significantly different ($p < 0.05$) from that of 3F (blue, $y = 1.0918x + 2.5027$, $R^2 = 0.9653$). (B) Venn diagram depicting probes of genes that have 2-fold up-regulation in response to butyrate treatment in 4F (B-up'ed, 4F) or 3F (B-up'ed, 3F) reprogramming. A total of 199 probes that are up-regulated by butyrate in 4F, but not 3F, reprogramming were analyzed further. (C) RT-qPCR verification of 3 randomly picked ES cell-enriched genes listed in Table S2-1, *Mreg*, *Krt12* and *Bex1*. Data presented is normalized to *Gapdh*.

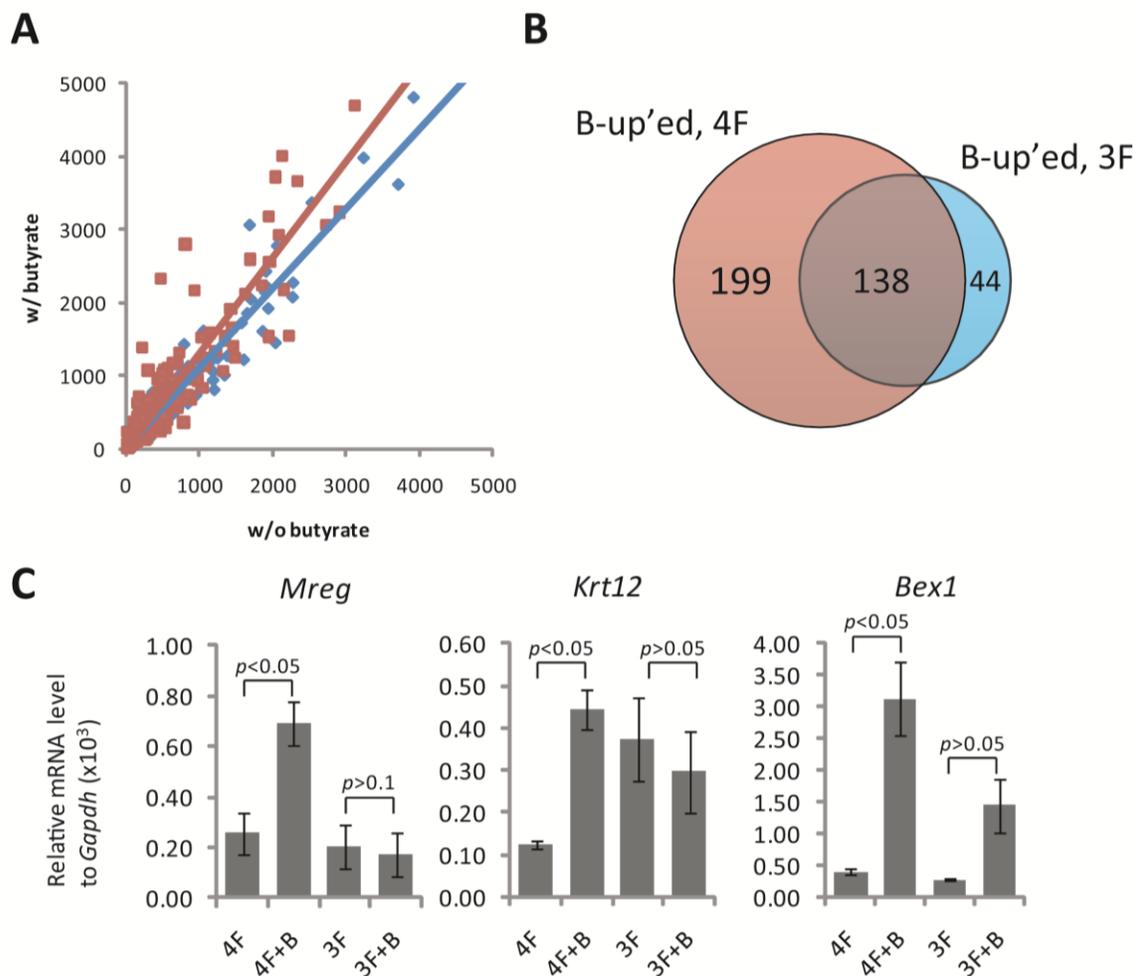


Figure S2-1 Expression of butyrate up-regulated genes in response to different concentration of butyrate.

Total RNA samples harvested 2 days after mock or butyrate treatment (0.25 mM, 0.5 mM and 1 mM) were subjected to RT-qPCR using primers specific for *Bex1*, *Mreg* and *Krt12*.

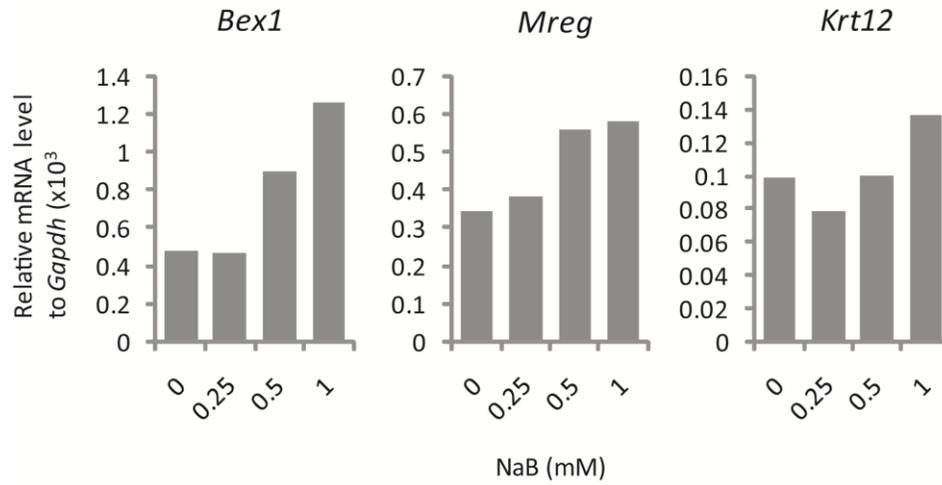


Figure S2-2 Butyrate treatment does not significantly suppress the p53-p21 pathway or Ink4a/Arf level.

(A) RT-qPCR analysis of the various samples with or without the treatment of butyrate indicates that butyrate treatment does not significantly suppress the p53-p21 pathway or Ink4a/Arf level in promoting the reprogramming. (B) Western blot analysis confirmed the RT-qPCR analysis presented in panel A, supporting the conclusion that butyrate treatment does not significantly suppress the p53-p21 pathway or Ink4a/Arf, excluding its involvement in mediating the butyrate effect.

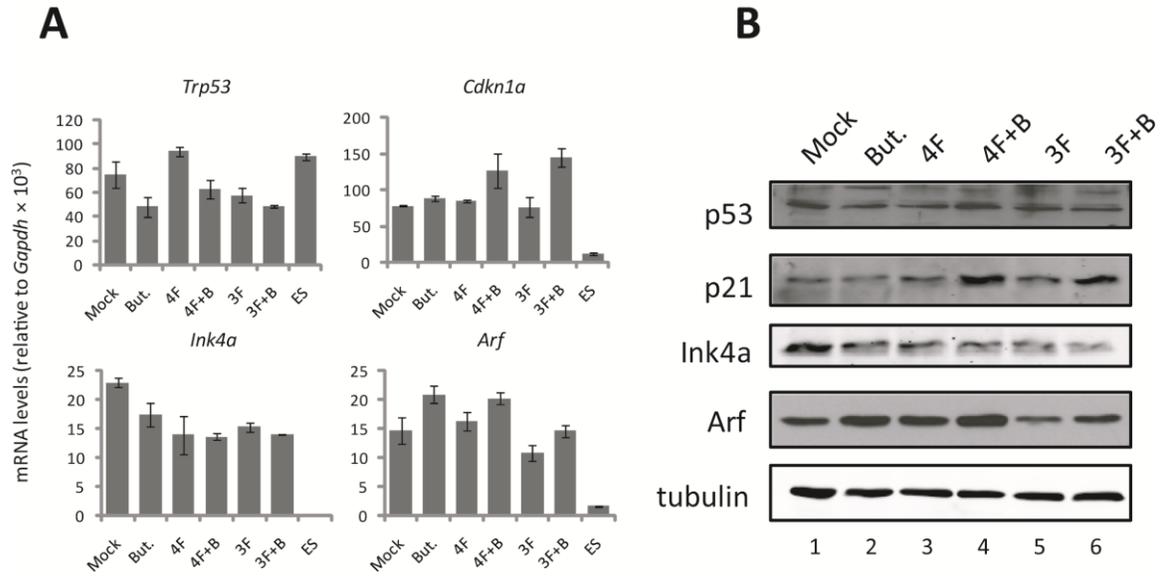


Table S2-1 Probes of ES cell-enriched genes that are 2-fold up-regulated by butyrate in 4F reprogramming, 3F reprogramming and both.

Probe ID	Symbol	Name
<i>Up-regulated only in 4F reprogramming</i>		
1432466_a_at	ApoE	apolipoprotein E
1418687_at	Arc	activity regulated cytoskeletal-associated protein
1448595_a_at	Bex1	brain expressed gene 1
1449887_at	Chmp4c	chromatin modifying protein 4C
1418709_at	Cox7a1	cytochrome c oxidase, subunit VIIa 1
1434170_at	Dcaf12l1	DDB1 and CUL4 associated factor 12-like 1
1416579_a_at	Epcam	epithelial cell adhesion molecule
1417883_at	Gstt2	glutathione S-transferase, theta 2
1449074_at	Kcnk4	potassium channel, subfamily K, member 4
1419230_at	Krt12	keratin 12
1457314_at	L1td1	LINE-1 type transposase domain containing 1
1416236_a_at	Mpzl2	myelin protein zero-like 2
1448265_x_at	Mpzl2	myelin protein zero-like 2
1437250_at	Mreg	melanoregulin
1429013_at	Mtap7d2	MAP7 domain containing 2
1438820_at	Rnf17	ring finger protein 17
1416627_at	Spint1	serine protease inhibitor, Kunitz type 1
1448562_at	Upp1	uridine phosphorylase 1
1431786_s_at	1190003J15Rik	RIKEN cDNA 1190003J15 gene
1417797_a_at	1810019J16Rik	RIKEN cDNA 1810019J16 gene

1455918_at		
<i>Up-regulated only in 3F reprogramming</i>		
1434073_at	Gprasp2	G protein-coupled receptor associated sorting protein 2
1426617_a_at	Ttyh1	tweety homolog 1 (Drosophila)
<i>Up-regulated in both 4F and 3F reprogramming</i>		
1428209_at	Bex4	brain expressed gene 4
1452004_at	Calca	calcitonin/calcitonin-related polypeptide, alpha
1442273_at	Ccdc158	coiled-coil domain containing 158
1417590_at	Cyp27a1	cytochrome P450, family 27, subfamily a, polypeptide 1
1460454_at	Glod5	glyoxalase domain containing 5
1430238_at	Got1l1	glutamic-oxaloacetic transaminase 1-like 1
1434670_at	Kif5a	kinesin family member 5A
1426255_at	Nefl	neurofilament, light polypeptide
1416965_at	Pcsk1n	proprotein convertase subtilisin/kexin type 1 inhibitor
1434292_at	Snhg11	small nucleolar RNA host gene 11 (non-protein coding)
1417426_at	Srgn	serglycin
1421606_a_at	Sult4a1	sulfotransferase family 4A, member 1
1419289_a_at	Syngn1	synaptogyrin 1
1418743_a_at	Tesc	tescalcin
1418744_s_at	Tesc	tescalcin
1424351_at	Wfdc2	WAP four-disulfide core domain 2

Table S2-2 Primers used for quantitative and semi-quantitative RT-PCR in Chapter Two.

Primer	Sequence
Oct4 F	CCAGAAGGGCAAAGATCAA
Oct4 R	GCTCCTGATCAACAGCATCA
Sox2 F	ACCAGCTCGCAGACCTACAT
Sox2 R	CCCTCCCAATTCCCTTGAT
Nanog F	AAGTACCTCAGCCTCCAGCA
Nanog R	GAAGTTATGGAGCGGAGCAG
Fbx15 F	ACATTGCCTCCCGACACTAC
Fbx15 R	GAAGGCAGGCAGATCTCAAG
Utf1 F	CGTCGCTACAAGTTCCTCAA
Utf1 R	CAGAGTGTCGGTGCTCGTAA
Gapdh F	CATGGCCTTCCGTGTTCTTA
Gapdh R	GCCTGCTTCACCACCTTCTT
Mreg F	GCGGCAGATCCGCAGGGAAG
Mreg R	CTCCCAGCTGGCGGGGAAGA
Krt12 F	GGCTCGCTGGCTGAAACCGA
Krt12 R	CTCCAGGCGAGCCTTGACGC
Bex1 F	TTCGGCAGCCCATCGCTCAC
Bex1 R	CGGGTCAGTGCTAACCGCCC

CHAPTER 3 KDM2B PROMOTES INDUCED PLURIPOTENT STEM CELL GENERATION BY FACILITATING GENE ACTIVATION EARLY IN REPROGRAMMING[#]

3.1 Introduction

Direct reprogramming from somatic cells to a pluripotent state can be achieved by introduction of defined transcription factors, such as Oct4, Sox2, Klf4 and c-Myc (Takahashi and Yamanaka 2006). The resultant iPS cells are molecularly and functionally similar to ES cells derived from the inner cell mass of a blastocyst (Hanna et al. 2010; Stadtfeld and Hochedlinger 2010). However, the process of iPS cell generation is highly inefficient, in terms of its frequency and the long latency prior to the establishment of pluripotency (Hanna et al. 2010; Stadtfeld and Hochedlinger 2010), due to some putative stochastic event(s) (Hanna et al. 2009; Yamanaka 2009). In the past several years, a number of factors that affect reprogramming efficiency through cell cycle dependent or independent mechanisms have been identified. For example, inhibition of the p53-p21 pathway and the *Ink4a/Arf* locus increases the reprogramming efficiency and accelerates the reprogramming dynamics by affecting cell proliferation (Banito et al. 2009; Hanna et al. 2009; Hong et al. 2009; Kawamura et al. 2009; Li et al. 2009; Utikal et al. 2009); whereas, ectopic expression of Nanog enhance reprogramming presumably through epigenetic mechanism without changing cellular proliferation status (Hanna et al. 2009).

[#] Adapted from Liang G, He J, Zhang Y. 2012. *Nat Cell Biol* doi:10.1038/ncb2483.

A recent study suggests that reprogramming from fibroblasts to iPS cells undergoes a series of transcriptional changes (Samavarchi-Tehrani et al. 2010). At the beginning, epithelial genes that alter the morphology of fibroblast to an ES cell-like state are first activated, followed by the activation of transcription factor Nanog and other pluripotent factors. After these waves of activation, mesenchymal genes are repressed, followed by the activation of “mature” pluripotent genes (Samavarchi-Tehrani et al. 2010). Consistent with these sequential molecular events, factors facilitating mesenchymal-to-epithelial transition (MET), such as BMPs, Tgf- β inhibitors and miR-200s, have been shown to be capable of promoting iPS cell generation (Ichida et al. 2009; Maherali and Hochedlinger 2009; Li et al. 2010; Samavarchi-Tehrani et al. 2010), indicating that activation of an transcription program during early stage of reprogramming is crucial for the establishment of pluripotency. However, how such a transcription program is activated is currently unknown.

Given that cell fate reprogramming is essentially reprogramming of epigenetic states (Goldberg et al. 2007), it is not surprising that chemical inhibitors of epigenetic enzymes (Huangfu et al. 2008a; Mikkelsen et al. 2008; Shi et al. 2008a; Liang et al. 2010), and certain chromatin-remodeling factors (Singhal et al. 2010) are able to promote iPS cell generation. Through studying epigenetic factors specifically enriched in ES cells, we found that Kdm2b (also named Jhdm1b and Fbx110), an H3K36me2-specific demethylase, is able to facilitate iPS cell generation. This property is not dependent on its effect on cell proliferation, but relies on its demethylase and DNA binding activities. Further analysis demonstrates that Kdm2b enhances reprogramming by promoting the

expression of epithelial genes, whose activation is required for achieving pluripotency (Chen et al. 2010; Li et al. 2010; Redmer et al. 2011).

3.2 Results

Kdm2b promotes iPS cell generation

To identify epigenetic factors that facilitate iPS cell generation, we focused on epigenetic factors enriched in ES cells. Kdm2b is one member of the Kdm2 protein family (Figure S3-1) capable of removing H3K36me2 (Tsukada et al. 2006; He et al. 2008). Kdm2b isoform 1 (IF1, hereafter referred as Kdm2b) is highly expressed in ES cells compared to mouse embryonic fibroblasts (MEFs) (Figure 3-1A) and is progressively up-regulated during reprogramming by Oct4, Sox2 and Klf4 (OSK) (Figure S3-2). To explore its potential role in iPS cell generation, we constructed a doxycycline-inducible lentiviral plasmid that expresses C-terminal Flag-tagged Kdm2b. Upon doxycyclin induction, Kdm2b is expressed about 50 times of the Kdm2b level in ES cells (Figure S3-3). When introduced into *Oct4-IRES-GFP/Rosa26-M2rtTA* MEFs together with the retroviral reprogramming factors OSK, it can increase Oct4-GFP⁺ colony numbers by 4–6 folds at days 12 and 16 of reprogramming (Figure 3-1B), indicating that Kdm2b is capable of enhancing OSK-mediated iPS cell generation.

We next asked whether Kdm2b is able to promote iPS cell generation in the presence of c-Myc by introducing Kdm2b into the cocktail of OSK plus c-Myc (OSKM). Results shown in Figure 3-1C demonstrate that Kdm2b is capable of increasing Oct4-GFP⁺ colonies numbers in the presence of c-Myc, indicating that Kdm2b likely facilitates iPS cell generation differently from c-Myc. By following the reprogramming

kinetics, we found that, although Kdm2b promotes iPS cell generation, it does not significantly shorten the latency time for the appearance of Oct4-GFP+ colonies, as the first Oct4-GFP+ colony appears around day 10 (OSK) or day 8 (OSKM) of reprogramming regardless whether Kdm2b is introduced to the reprogramming system (Figure 3-1D). This observation differs from cell cycle-based enhancement of iPS cell generation, including depletion of p53 and p21 (Hanna et al. 2009; Hong et al. 2009; Kawamura et al. 2009) and inhibition of *Ink4a/Arf* (Li et al. 2009; Utikal et al. 2009), which significantly shortens the latency time of iPS cell generation through promoting cell proliferation (Hanna et al. 2009). Therefore, Kdm2b enhances reprogramming likely by using a cell cycle-independent mechanism similar to Nanog overexpression (Hanna et al. 2009).

To further confirm the role of Kdm2b in reprogramming, we introduced small hairpin RNA (shRNA) (He et al. 2008), which depletes the *Kdm2b* mRNA level by around 80% (Figure 3-1E), to the reprogramming cells co-transduced with OSK. We found that the reprogramming efficiency is reduced to half when shRNA against Kdm2b (Kdm2b-i) is introduced (Figure 3-1F), indicating that Kdm2b is required for optimal induction of iPS cells. Collectively, the above studies demonstrate that Kdm2b facilitates iPS cell generation.

iPS cells generated in the presence of Kdm2b are pluripotent

Next, we set out to characterize the Oct4-GFP+ iPS cells generated with OSK in the presence of exogenous Kdm2b. After 16–18 days of iPS cell induction, Oct4-GFP+ colonies were manually picked up and propagated into cell lines in the absence of

doxycyclin. The resulting cell lines exhibit typical ES cell morphology with Oct4-GFP expression (Figure 3-2A). These cells are positive for alkaline phosphatase (AP) activity and express ES cell markers SSEA-1, as well as pluripotent transcription factors Nanog and Sox2 (Figure 3-2B). RT-qPCR demonstrates activation of endogenous Oct4, Sox2 and Nanog genes to a level similar to that in ES cells (Figure 3-2C). In contrast, the transgenes, including the retrovirus introduced Oct4, Sox2 and Klf4, and the lentiviral inducible Kdm2b, are mostly silenced in the iPS cell lines, as shown by semi-quantitative RT-PCR (Figure 3-2D). When implanted into immunodeficient mice, these iPS cells generated teratomas with tissues belonging to all three germ layers (Figure 3-2E). Importantly, these iPS cells are competent for chimera generation when injected into blastocysts (Figure 3-2F). These data support that the iPS cells generated in the presence of exogenous Kdm2b are pluripotent.

Kdm2b facilitates iPS generation in an enzymatic activity-dependent and cell proliferation alteration-independent manner

To understand how Kdm2b facilitates iPS generation, we attempted to determine the domains of Kdm2b important for this property. To this end, we carried out reprogramming with point mutations in the catalytic JmjC domain (Tsukada et al. 2006; He et al. 2008) and the DNA binding zinc finger-CXXC (ZF) domain (Voo et al. 2000; Blackledge et al. 2010). We first confirmed that both the mutants and wild-type (WT) Kdm2b are expressed at a similar level upon doxycyclin induction (Figure 3-3a). Consistent with Kdm2b's enzymatic activity, doxycycline induced expression of WT and ZF mutant, but not the catalytic mutant, leads to a specific decrease of H3K36me2 levels (Figure 3-3A). Importantly, these three forms of Kdm2b display differential capacity in

enhancing iPS cell generation when introduced to the reprogramming system. Mutations on either JmjC or ZF abrogate the capacity of Kdm2b in promoting iPS cell generation (Figure 3-3B), indicating that both the demethylase activity and the DNA binding activity of Kdm2b are functionally important for its role in iPS cell generation.

Previous studies have demonstrated a role for Kdm2b in promoting cell proliferation by repressing senescence (He et al. 2008; Pfau et al. 2008; Tzatsos et al. 2009). To determine whether the capacity of Kdm2b in promoting iPS cell generation is solely mediated by promoting cell proliferation, we examined how ectopic expressions of Kdm2b affect cell proliferation in the context of OSK-mediated reprogramming. We found that although ZF domain of Kdm2b is critical for its role in promoting iPS generation, it is not required for its ability to promote cell proliferation as both the WT and the ZF mutant promoted cell proliferation to a similar level (Figure 3-3C). These data suggest that the ability of Kdm2b in enhancing cell proliferation is not sufficient for promoting iPS cell generation. To illustrate that the cell proliferation effect of Kdm2b is not a major contributor to its effect on reprogramming, we normalized the reprogramming efficiency by dividing the Oct4-GFP⁺ colony numbers with the total cell numbers in the reprogramming cell populations. Results presented in Figure 3-3D indicate that the presence of Kdm2b still increases reprogramming efficiency for more than 4 folds. Although the JmjC mutant also increases the reprogramming efficiency, it is only about half of the effects exhibited by the WT, while the ZF mutant completely loses the effect on enhancement. These analyses indicate that the demethylase activity and the DNA binding capacity of Kdm2b are essential for its function in promoting iPS cell

generation, and that cell proliferation promoted by Kdm2b is not a major contributing factor for its role in promoting reprogramming.

To further ascertain that Kdm2b-mediated suppression of senescence is not a major contributor to its role in reprogramming enhancement, we examine the transcript levels of *Ink4a*, *Arf* and *Ink4b* during Kdm2b-mediated reprogramming. We found that, *Ink4a*, but not *Arf* or *Ink4b*, is significantly downregulated by Kdm2b in the first 12 days of OSK reprogramming (Figure 3-4A). However, the protein level of *Ink4a* is less affected by Kdm2b, probably due to the stability of this protein (Figure 3-4B). To examine whether repression of this locus contributes to Kdm2b-mediated enhancement of reprogramming, we introduced shRNAs, that depleted *Ink4a* and *Arf* by more than 90% at mRNA and protein levels (Figure 3-4C, -D), to the reprogramming cells transduced with OSK and OSK plus Kdm2b. We found that, albeit the deficiency of *Ink4a/Arf*, Kdm2b is still able to increase Oct4-GFP⁺ colony numbers by 3–4 folds compared to OSK reprogramming, suggesting that the Kdm2b-mediated enhancement of iPS cell generation is largely independent of its role in downregulating *Ink4a/Arf* (Figure 3-4E). Consistent with previous finding that suppression of the senescence pathway generally promotes reprogramming (Banito et al. 2009; Li et al. 2009; Utikal et al. 2009), we also observed a great enhancement of reprogramming efficiency when *Ink4a/Arf* are depleted regardless whether Kdm2b is used (Figure 3-4E). Collectively, our data suggest that the ability of Kdm2b to suppress cellular senescence is not the major reason for its role in enhancing reprogramming.

Kdm2b functions early during the reprogramming process

Given that the role of Kdm2b in suppressing senescence and/or promoting cell proliferation is not a major contributor, and that introduction of Kdm2b does not enhance expression of transduced reprogramming factors (Figure S3-4), we searched for alternative mechanisms. Previous studies suggest that reprogramming can be divided into stages with distinct molecular features (Brambrink et al. 2008; Stadtfeld et al. 2008; Samavarchi-Tehrani et al. 2010), and factors that promote iPS cell generation can function at different stages during reprogramming (Ichida et al. 2009; Silva et al. 2009; Sridharan et al. 2009a). Therefore, we attempted to first determine the time window at which Kdm2b enhances iPS generation. To this end, we added doxycyclin to induce the expression of Kdm2b for different durations and analyzed its effect on the efficiency of iPS cell generation. Compared with the OSK control, enforced expression of Kdm2b in the first 8 days constantly enhanced the reprogramming efficiency, and prolonged induction of Kdm2b does not further enhance iPS cell generation (Figure 3-5A, upper panels). As a complement, we also induced Kdm2b expression beginning at different days after the initiation of reprogramming and found that the increase in reprogramming efficiency is largely proportional to the length of the doxycyclin treatment (Figure 3-5A, lower panels). Together, these studies indicate that Kdm2b promotes iPS cell generation by functioning early during the reprogramming process.

Kdm2b amplifies early gene activation during reprogramming

To understand the effect of Kdm2b during reprogramming at the molecular level, we performed microarray studies on cells harvested at days 4, 8 and 12 during the OSK

reprogramming in the presence or absence of Kdm2b. This analysis revealed a total of 418 probes that showed a minimum of 2-fold upregulation for at least one of the three selected time points, while 143 probes showed at least 2-fold downregulation (Figure 3-5B). Hierarchical clustering of these probes revealed that the Kdm2b upregulated genes can be grouped into three distinct clusters (cluster I, II and III) based on the time of their activation. Cluster I consists of probes upregulated at days 8 and 12; cluster II consists of probes upregulated from days 4 to 12; and cluster III probes are only upregulated on day 4 (Figure 3-5B). In contrast, Kdm2b downregulated genes are distributed more ambiguously in terms of their affected timing.

Gene ontology analysis revealed that Kdm2b affected genes are exceptionally enriched for adhesion processes (p -value = $10^{-10} \sim 10^{-12}$). Other terms with a significant p -value ($< 10^{-3}$) includes those related to cell morphology, development and epithelium-related processes (Figure 3-5B). Of particular relevance, recent studies have demonstrated that reprogramming starts from activation of epithelial adhesion genes, which results in dramatic change in cell morphology and adhesion properties (Li et al. 2010; Samavarchi-Tehrani et al. 2010). Consistent with the notion that Kdm2b functions early at the beginning of reprogramming, Kdm2b likely plays a role in regulating these early responsive genes during reprogramming. To explore this possibility, we compared the Kdm2b-affected genes with a list of “signature” genes (Figure S3-5) whose expressions are dynamically regulated at different stages of the reprogramming process (Samavarchi-Tehrani et al. 2010). We found that Kdm2b-affected genes include some early activated “signature” genes, such as the epithelial genes *Cdh1*, *Cldn3*, *-4* and *-7*, *Epcam*, *Esrp1* and *Ocln*, as well as pluripotent genes *Nanog*, *Dppa5a* and *Tdgf1* (Figure

3-5D). Indeed all the previously identified early activated genes (Samavarchi-Tehrani et al. 2010) have enhanced expression in the presence of Kdm2b, while the late activated genes or the repressed genes are not significantly altered by Kdm2b (Figure S3-5). Therefore, Kdm2b appears to be involved in the activation of the early responsive genes during reprogramming.

Given that Kdm2b exerts its effect till day 8 during reprogramming (Figure 3-5A), the gene expression changes observed at days 4 and 8 are likely to be direct or primary effect of Kdm2b, while those changed at days 8 and 12 are more likely due to secondary effect. To better understand these changes, we grouped the Kdm2b-affected genes based on timing of the effect and found that most of the upregulated genes at days 4 and 8 continue to be upregulated at later time points (Figure 3-5E), while genes upregulated at days 8 and/or 12 are mostly not shared with the sets of genes activated at an earlier time points. Similarly, most downregulated genes at day 4 continue to be downregulated at later time points, while those downregulated at days 8 and 12 are mostly not shared with those downregulated at day 4 (Figure 3-5E). Such a gene distribution pattern prompts us to hypothesize that introduction of Kdm2b amplifies a putative reprogramming transcriptional cascade (see discussion).

To probe into the potential difference among genes activated by Kdm2b at different time during reprogramming, we carried out gene ontology analysis of the different sections of genes in Figure 3-5E. We found that genes upregulated from days 4 through 12 are enriched in adhesion molecules (Figure 3-5F), consistent with the notion that Kdm2b functions from the beginning in regulating cell adhesion during reprogramming. Interestingly, genes that are significantly upregulated at day 4 only are

enriched in regulation of adhesion; those upregulated at days 8 and 12 are enriched for developmental genes in addition to genes of adhesion molecules; and those only upregulated at day 12 are more enriched for developmental process (Figure 3-5f). Such changes in gene function implicates that a developmental-related transcription program is triggered sometime after an initial adhesion-related program during reprogramming. Based on the delayed onset of the activation of development-related transcription program, it is likely that its activation requires the complement of preceding transcription activation, at least, to a certain extent, hence very likely constituting a transcription cascade. Later we found that the timing of the switch in gene functions coincides with that of *Nanog* activation (Figure 3-6A). For the downregulated gene, only those downregulated at day 8 showed an enriched functionality in immune response (Figure 3-5E), whose significance remains to be clarified.

Next, we performed RT-qPCR to verify the upregulation of genes encoding for epithelial cell markers, such as *Cdh1* (also known as E-cadherin), *Crb3* and *Epcam*, as well as desmosomes components *Dsg2* and *Dsp* (Figure 3-6A, -B). Enhanced expression of these genes starts at day 4 (Figure 3-6B), and more obviously affected at days 8 and 12 (Figure 3-6A, -C). We also note that genes activated earliest at day 4 include those encodes transcription factors with a strong indication functioning in adhesion and development, including *Irf6*, which is important for cleft palate (Richardson et al. 2009; Thomason et al. 2010), and *Insm1*, which plays a role in pancreatic and neuronal development (Gierl et al. 2006; Mellitzer et al. 2006; Farkas et al. 2008). Interestingly, these genes are also upregulated by OSK and introduction of *Kdm2b* further augments their activations (Figure 3-6A, -B). The function of these early responsive genes in

reprogramming remains to be characterized. Furthermore, the expression levels of pluripotent genes, such as *Nanog* and *Tdgf1*, are also enhanced by Kdm2b at days 8 and 12 of induction, but not at day 4 (Figure 3-6A, -B). Since Kdm2b functions at the initial 8 days of OSK reprogramming (Figure 3-5A), these pluripotent genes are likely to be regulated by Kdm2b indirectly.

In addition to adhesion and pluripotent genes, we also examined the expression of mesenchymal-specific transcription factors, which has been previously shown to be downregulated following the activation of *Nanog* and other pluripotent genes (Samavarchi-Tehrani et al. 2010). We observed a mild downregulation of *Snai1*, *Snai2*, *Zeb1* and *Zeb2* at days 8 and/or day 12 (Figure S3-5A) although their downregulation is less than 2 folds and consequently not picked up by the microarray analysis (Figure 3-5D). Finally, we did not observe activation of endogenous *Oct4* and *Sox2* during the first 12 days of reprogramming (Figure S3-5B), consistent with the fact that even in the case of Kdm2b-assisted reprogramming, only a very small fraction of the starting cells gained pluripotency in the 12 days of reprogramming monitored (Figure 3-1B) and hence their existence could not be reflected at mRNA level in the bulk population of reprogramming cells. Collectively, data presented above suggest that Kdm2b facilitates activation of genes related to epithelial adhesion, which may in turn activate downstream genes including pluripotent genes. Thus, it seems sensible that Kdm2b promotes iPS cell generation by facilitating the initiation of a putative transcription cascade.

Since mutations in JmjC domain and ZF domain abrogate the capacity of Kdm2b in promoting iPS cell generation, we asked whether these mutations affect the capacity of Kdm2b to argument the activation of these genes. RT-qPCR and Western blot analysis

demonstrate that both mutations abolish the ability of Kdm2b to activate *Cdh1* and *Epcam* (Figure 3-6D, -E), as well as subsequent activation of *Nanog* (Figure 3-6D). These results support a link between the ability of Kdm2b in activating these genes and its role in facilitating iPS cell generation, consistent with the notion that Kdm2b enhances iPS cell generation by facilitating early gene activation.

Kdm2b acts in concert with the key reprogramming factors in early gene activation

To understand how Kdm2b facilitates activation of early reprogramming genes, we asked whether Kdm2b activates these genes alone or in concert with the key reprogramming factors. To this end, we introduced individual factors and different factor combinations into MEFs and examined their effects on the expression of early responsive epithelial genes, such as *Cdh1*, *Crb3* and *Epcam*, as well as the later activated *Nanog*. We found that, overall when any of the OSK factor is omitted, activation of these genes is either greatly compromised or completely abolished even in the presence of Kdm2b (Figure 3-6F). This observation is consistent with the fact that Kdm2b cannot replace any of the OSK in iPS cell generation (data not shown). However, we note that Kdm2b alone does exhibit a 5-fold activation on *Cdh1* gene when compared to non-transduced MEF cells. Nevertheless, this activation appears to be minor (5%) when compared to the activation by OSK (Figure 3-6F). Similar to previous report (Li et al. 2010), transduction of *Klf4* alone can partially activate *Cdh1* compared to OSK transduction. Its activation can be further boosted when Kdm2b is combined with *Klf4*. (Figure 3-6F). For *Crb3*, *Epcam* and *Nanog*, Kdm2b barely activate them in the absence of any the OSK (Figure 3-

6F). Based on these results, we conclude that Kdm2b amplifies gene activation through cooperating with the key reprogramming factors OSK.

Kdm2b binds to and regulates the H3K36me2 level of the promoter of early activated genes

To determine whether Kdm2b directly contributes to the activation of the early responsive genes, we asked whether Kdm2b binds to these genes. Taking advantage of the Flag epitope on the Kdm2b constructs, we performed chromatin immunoprecipitation (ChIP) analysis using cells at day 4 of reprogramming. This analysis demonstrates that Flag-Kdm2b is enriched in the promoter region of the early activated genes, including *Cdh1*, *Epcam*, *Dsg2*, *Dsp* and *Irf6*; however, no enrichment is detected at the *Nanog* promoter, consistent with the notion that Kdm2b contributes to Nanog activation indirectly (Figure 3-7A). We also profiled the localization of exogenous Kdm2b at the *Cdh1* locus, finding that Kdm2b localizes from the promoter to the middle of the gene body, but not in the region proximal to 3' end or upstream beyond the promoter (Figure 3-7B, C). Parallel ChIP experiments demonstrate that the level of H3K36me2 at the promoter of *Cdh1*, *Dsp* and *Irf6* is decreased upon the introduction of Kdm2b (Figure 3-7D), but H3K4me2, an irrelevant modification to Kdm2b enzymatic activity, remains unchanged, consistent with the fact that Kdm2b preferentially removes H3K36me2 (He et al. 2008). These results support the notion that Kdm2b contributes to the activation of early responsive genes by binding to and demethylating H3K36me2 on their promoters.

Enhanced early gene activation mediates the effect of Kdm2b in reprogramming

Epithelial genes, which are activated early (Li et al. 2010; Samavarchi-Tehrani et al. 2010) and whose activation is enhanced by Kdm2b during reprogramming, are subjected to repression by Tgf- β signaling (Heldin et al. 2009). To demonstrate that increased expression of these genes contributes to the increased reprogramming efficiency by Kdm2b, we perform reprogramming by OSK plus Kdm2b in the presence of Tgf- β . We found that, upregulation of epithelial genes, such as *Cdh1* and *Epcam*, by Kdm2b are abrogated upon Tgf- β treatment (Figure 3-8A). Importantly, Kdm2b-enhanced iPS generation is also abrogated by Tgf- β (Figure 3-8B), consistent with the notion that Kdm2b-enhanced expression of the epithelial genes contributes to its effect on reprogramming. To directly address the role of these epithelial genes in mediating Kdm2b-enhanced reprogramming, we focused on *Cdh1*, one of the epithelial genes directly regulated by Kdm2b. We generated three shRNAs that efficiently deplete *Cdh1* (Figure 3-8C, -D) and asked whether the effect of Kdm2b-enhanced OSK reprogramming is affected by these shRNAs. Results shown in Figure 3-7E demonstrate that Kdm2b-mediated enhancement of reprogramming is reduced by 60 – 90% upon *Cdh1* depletion, suggesting that *Cdh1* is one of the key downstream targets that mediating Kdm2b's effect in reprogramming. We also tested whether enforced expression of *Cdh1* can enhance OSK-mediated reprogramming; however, consistent with previous reports (Li et al. 2010; Redmer et al. 2011), we found that adding *Cdh1* into the reprogramming factor cocktails failed to enhance iPS cell generation, indicating that overexpressing *Cdh1* alone is not sufficient to mimic the effect of Kdm2b. Nevertheless, our data demonstrates that *Cdh1*, at least, in part, mediates Kdm2b's effect on reprogramming.

3.3 Discussion

Kdm2b promotes reprogramming independent of its effect on cell proliferation

Despite cellular senescence is a “roadblock” for reprogramming (Li et al. 2009; Utikal et al. 2009), and that Kdm2b is capable of suppressing senescence (He et al. 2008; Tzatsos et al. 2009), several lines of evidence suggest that Kdm2b promotes iPS generation is largely independent of its role in senescence and/or cell proliferation. First, although Kdm2b mildly increases cell proliferation during reprogramming (Figure 3-3C), when the proliferation effect is subtracted, Kdm2b still exhibits more than 4-fold increase in reprogramming efficiency (Figure 3-3D). Second, although an intact ZF domain is required for Kdm2b to promote iPS cell generation, mutation on this domain does not alter its ability to stimulate cell proliferation, suggesting that the roles of Kdm2b in promoting proliferation and iPS cell generation are independent and separable. Third, when *Ink4a/Arf*, the key senescence regulators and documented targets of Kdm2b (Tzatsos et al. 2009), are depleted, Kdm2b is still capable of promoting iPS generation (Figure 3-4E), indicating that the ability of Kdm2b in antagonizing senescence is not a major contributor for its role in promoting iPS cell generation. Lastly, Kdm2b does not significantly shorten the latency time from induction to the appearance of first Oct4-GFP⁺ colony (Figure 3-1d), which is consistent with the reprogramming kinetics of cell cycle-independent enhancement of reprogramming (Hanna et al. 2009).

While our manuscript is under review, Pei and colleagues published data demonstrating that Kdm2b (Jhdm1b in their report) greatly enhances iPS cell generation in the presence of vitamin C and suggest that the effect of Kdm2b is mediated through

promoting cell cycle progression and activation of microRNA cluster 302/367 (Wang et al. 2011). However, most of the mechanistic studies are performed in the presence of both vitamin C, and therefore, the specific contribution from Kdm2b to iPS cell generation is not addressed. Indeed, the same group previously found that treatment of vitamin C is capable of alleviating the blockage of cell cycle progression imposed by p53-p21 (Esteban et al. 2010). Since vitamin C has more pronounced effect on iPS cell generation than Kdm2b (Wang et al. 2011), it is possible that the effect of vitamin C on cell cycle progression might mask the effect of Kdm2b and consequently the contribution of Kdm2b to iPS cell generation might be overshadowed.

A transcription cascade model for gene regulation in reprogramming

By analyzing the gene expression changes within or immediately after the functioning time window of Kdm2b, we found that exogenous Kdm2b enhances the expression of a set of early activated genes during reprogramming (Figure 3-5D and Figure S3-4). The enhanced expression of epithelial genes (Cdh1, Epcam and etc.) and other uncharacterized genes (Irf6 and Insm1) takes place in the functioning time window of Kdm2b (Figure 3-5A, -6A, -6B). Following this first wave of activation, the expression of Nanog and other pluripotency factors are upregulated (Figure 3-6A, B) concomitant with enrichment of developmental genes in the upregulated gene group (Figure 3-5F). Meanwhile, some mesenchymal genes, whose downregulation follows Nanog activation (Samavarchi-Tehrani et al. 2010), begin to display magnified downregulation in the presence of Kdm2b (Figure S3-5A). Soon after Nanog activation (day 8; Figure 3-6A), first Oct4-GFP positive colonies are observed at day 10 (Figure 3-1D). The observation that Kdm2b amplifies these sequential transcription events

prompted us to propose a transcription cascade model to explain how Kdm2b might contribute to reprogramming (Figure 3-8F). We propose that Kdm2b facilitates initial gene activation that occurs on the epithelial genes, causing an amplified transcription cascade, which in turn enhances the activation of pluripotent genes such as Nanog, eventually resulting in an increase in reprogramming efficiency (Figure 3-8F).

This model is consistent with previous observation that Tgf- β inhibitors induce Nanog expression and enhance reprogramming (Ichida et al. 2009; Maherali and Hochedlinger 2009) as Tgf- β signaling inhibits epithelial gene expression (Heldin et al. 2009; Li et al. 2010) and suppression on Tgf- β potentially helps activation of epithelial genes. It also agrees with the notion that Nanog is not required for initiating reprogramming but plays a key role in driving the “pre-iPS” cells to pluripotency (Silva et al. 2009). Uncovering the potential links between transcription events is crucial for attesting the transcription cascade model.

Kdm2b facilitates early gene activation in reprogramming

Our study revealed that Kdm2b promotes activation of early responsive genes at the beginning of the reprogramming process (Figure 3-5, 6). Although the details of how Kdm2b contributes to the activation of these genes need to be revealed, our studies provide a few clues. First, upregulation of these early-activated epithelial genes is not due to a decrease in the levels of mesenchymal transcription factors as downregulation of mesenchymal transcription factors takes place after the activation of epithelial genes (Figure S3-5A). Secondly, we found that the ability of Kdm2b in promoting the activation of early reprogramming genes is dependent on the presence of OSK in the

reprogramming cocktail (Figure 3-6F). This is consistent with our observation that Kdm2b cannot replace any of the OSK for iPS cell generation under our reprogramming conditions (data not shown). Finally, we demonstrate that Kdm2b binds to the promoter of some of the activated epithelial genes and maintains H3K36me2 at a lower level (Figure 3-7A, D) supporting a direct role for Kdm2b in regulating expression of these genes. How Kdm2b is recruited to its targets and how Kdm2b promotes the activation of epithelial genes remains to be determined.

H3K36 methylation has been mostly linked to transcriptional elongation and the studies are mostly carried out in yeast. Therefore, the role of promoter H3K36 methylation in mammalian cells is largely unknown. A recent genome-wide study in mouse ES cells revealed that Kdm2a, a paralog of Kdm2b, binds to nonmethyl-CpG island promoters and depletes H3K36me2 at these promoters (Blackledge et al. 2010). Given that Kdm2b also contains a ZF (CXXC) domain, a nonmethyl-CpG binding domain, we anticipate that Kdm2b should also localize to the nonmethyl-CpG islands. In addition to devoid of H3K36me2, nonmethyl-CpG island promoters are also enriched for H3K4me3 installed by Cfp1-mediated recruitment of Setd1 (Thomson et al. 2010), as well as the 5-methylcytosine oxidase Tet1 (Wu et al. 2011b). Since all these chromatin features are linked to gene activation, it is not surprising that Kdm2b is involved in gene activation in our study. It is possible that binding of Kdm2b to nonmethyl-CpG promoters eliminates H3K36me2, which may facilitate OSK-mediated gene activation by establishing a platform for cofactor recruitment. Such a scenario explains why Kdm2b can promote gene activation (Figure 3-5, 6) and facilitate OSK-directed reprogramming (Figure 3-1) but fails to substitute for any of OSK (data not shown). Future studies should

reveal the molecular mechanisms underlying how demethylation of H3K36me2 contributes to gene activation.

3.4 Materials and methods

Plasmids and virus preparation

Mouse Kdm2b (isoform 1) were amplified from cDNA, fused with a C-terminal Flag tag, and cloned into a doxycyclin inducible lentiviral vector pTYF-TRE. The JmjC (H211A, D213A) and ZF (C573A C576A C579A) mutants were constructed by mutagenesis PCR and confirmed by sequencing. Target sequences of control shRNA and shRNAs against Kdm2b (He et al. 2008), Ink4a/Arf and Cdh1 are described in Table S3-1. These shRNAs were expressed in lentiviral plasmid pTY-U6-Pgk-Puro. Lentivirus was prepared by cotransfection of pTY/pTYF plasmids with pHP, pHEF1 α -VSVG and pCEP4-Tat into 293T cells, and harvested at 24, 36 and 48 hours after transfection. Viral supernatant was filtered through 0.45- μ m membrane and concentrated by spin column before applied to MEFs. Retroviral plasmids pMXs-*Oct4*, *Sox2*, *Klf4* and *c-Myc* were obtained from Addgene, and retrovirus was prepared as previously described (Liang et al. 2010).

MEF derivation and iPS cell generation

MEFs for iPS cell generation were prepared from E13.5 embryos of *Oct4-IRES-GFP/Rosa26-M2rtTA* double knock-in mice. To derive iPS cells, MEFs at the first 2 passages were seeded onto 6-well plates at a density of 1×10^5 cells per well, 16 h before viral infection. Two doses of retrovirus and/or one dose of lentivirus was applied within

48 hours in the presence of polybrene (10 µg/ml). 24 hours after the second retroviral transduction, the virus supernatant was withdrawn and the day was designated as day 0 post-transduction. Subsequently, iPS cells were induced for 12–18 days in mouse ES cell medium (DMEM with 15% FBS, non-essential amino acid, GlutaMax, sodium pyruvate, β-mercaptoethanol, penicillin/streptomycin and 1,000 U/ml leukemia inhibitory factors) in the presence of doxycyclin (1 µg/ml). If indicated, Tgf-β treatment was carried out by applying Tgf-β1 (R&D Systems) at 2 ng/ml. Oct4-GFP⁺ colonies were counted on selected days from day 6 to day 18. Reprogramming efficiency was presented as the number of Oct4-GFP⁺ colonies derived from 1×10⁵ MEFs. Relative reprogramming efficiency over a control induction is also used in some cases. At day 18, Oct4-GFP⁺ colonies were manually picked, trypsinized and seeded onto mitomycin C-treated feeder MEF cells. The derived iPS cell lines were propagated in mouse ES cell medium in the absence of doxycyclin for at least 8 passages before being characterized.

Cell staining, teratoma assay and chimera generation

For immunofluorescent staining, antibodies against SSEA-1 (Chemicon mAB4301, clone MC-480), Nanog (Bentyl, IHC-00205) and Sox2 (Millipore, AB5603) are applied at a concentration of 1:500, 1:250 and 1:1000, respectively. AP staining was carried out with AP detection kit (Millipore). Teratoma analysis was performed as previously reported (Liang et al. 2010). For generation of chimeric mice, twelve week old Albino B6 (C57Bl/6J-Tyr^{c-2J}) female mice were stimulated to superovulation by injected with pregnant mare serum gonadotropin (2.5 IU) followed by administration of human chorionic gonadotropin (5 IU) 47 hours later. The female mice were subsequently mated with Albino B6 stud males and blastocysts were harvested on gestation day 3.5.

On the day of microinjection, iPS cells (line 8 and 17) were rinsed twice with PBS, dissociated with 0.05% trypsin, washed once with and then resuspended in Knockout DMEM supplemented with 15% FBS. Each blastocyst was injected with 10–15 iPS cells using a piezo impact micromanipulator. Injected embryos were then implanted into the uterus of pseudopregnant Swiss Webster recipient females.

RT-PCR and Western blotting

Quantitative and semi-quantitative RT-PCR was carried out using primers in Table S3-2. Western blotting was performed using antibodies against Arf (Santa Cruz sc-32748, clone 5-C3-1, 1:200), Cdh1 (Cell Signal 3195, clone 24E10, 1:1000), Epcam (Abcam ab32392, clone E144, 1:500), Flag (Sigma F1804, clone M2, 1:5000), histone H3 (Abcam ab1791, 1:5000), H3K4me2 (Active Motif 39141, 1: 1000), H3K36me1 (Abcam ab9048, 1:1000), H3K36me2 (Tsukada et al. 2006) (1:1000), H3K36me3 (Abcam ab9050, 1: 1000), Ink4a (Santa Cruz sc-1207, 1:200) and α -tubulin (Sigma T6199, clone DM1A, 1:2000).

Microarray analysis

RNA samples were extracted from cells transduced with OSK or OSK plus Kdm2b at post-transduction day 4, 8 and 12. The reverse transcription and hybridization procedure was carried out as previously described (Liang et al. 2010). The microarray data were analyzed with GeneSpring software and are available in ArrayExpress database (<http://www.ebi.ac.uk/arrayexpress/>) with accession number E-MEXP-3433.

Chromatin Immunoprecipitation (ChIP)

Cells transduced with OSK plus Flag-tagged Kdm2b were harvested for ChIP at post-transduction day 4. ChIP was performed using Imprint ChIP Kit (Sigma) according to the manufacturer's instruction. Chromatin was prepared by sonication at 4 °C on Bioruptor 300 (Diagenode) with high magnitude for 10 cycles with 30 seconds on and 30 seconds off. For each precipitation reaction, chromatin from 2×10^5 cells was applied to a Stripwell pre-bound with antibodies against Flag (Sigma F1804, clone M2), H3K36me2 (Tsukada et al. 2006), H3K4me2 (Active Motif 39141) or mouse IgG. If necessary, immunoprecipitated and purified DNA fragments were subjected to amplification using Whole Genome Amplification Kit (Sigma). Immunoprecipitated or amplified DNA was analyzed by quantitative PCR using primers listed in Table S3-3.

Standard Error Reporting

Data presented with error bars were collected from samples from at least two independent preparations. The number of preparations (n) for each experiment is denoted in the figure legend. Error bars reported represent standard error of the mean.

Figure 3-1 Kdm2b promotes iPS cell generation.

(A) Isoform 1 (IF1) of *Kdm2b* is highly expressed in mouse ES cells. RT-qPCR analysis of the expression levels of *Kdm2a* and different isoforms of *Kdm2b* (*IF1*, *IF2* and *IF3*) in mouse ES cells (ESC) and embryonic fibroblasts (MEF). Expression levels are normalized to *Gapdh* and compared to the expression level in MEF. $n = 2$. (B) Kdm2b increases the efficiency of iPS cell generation when co-introduced with Oct4, Sox2 and Klf4 (OSK). The efficiency is represented by the number of Oct4-GFP⁺ colony counted at post-transduction days 12 and 16 from 1×10^5 starting MEFs. $n = 3$. (C) Kdm2b increases the reprogramming efficiency in the presence of c-Myc. Shown are numbers of Oct4-GFP⁺ colony at day 12 reprogrammed by OSK plus c-Myc (OSKM) in the presence or absence of Kdm2b. $n = 3$. (D) The reprogramming kinetics from day 6 to day 16 using different combinations of reprogramming factors. $n = 3$. (E) RT-qPCR analysis of the knock-down efficiency of *Kdm2b*. The results were normalized to the *Gapdh* level and shown as relative to the control shRNA treatment (Control-i). $n = 2$. (F) Knockdown of Kdm2b reduced OSK-mediated reprogramming efficiency. Oct4-GFP⁺ colony numbers at days 14 and 18 are shown for OSK reprogramming with control (Control-i) or Kdm2b knockdown (Kdm2b-i). $n = 2$.

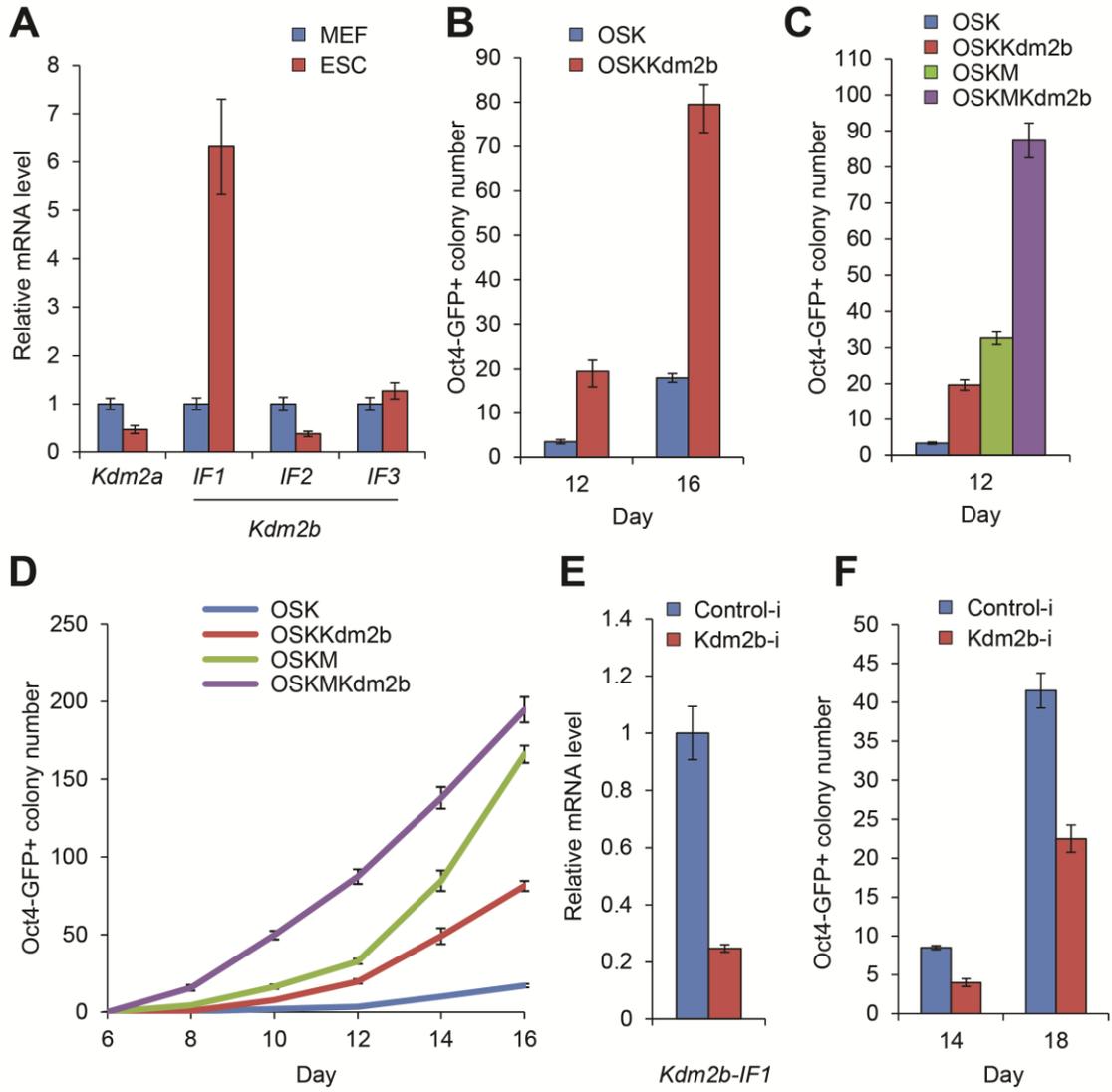


Figure 3-2 iPS cells generated by enforced expression of OSK plus Kdm2b are pluripotent.

(A) The morphology and Oct4-GFP expression of a representative iPS cell line derived from Kdm2b-assisted reprogramming. Scale bar, 200 μm . (B) Alkaline phosphatase (AP) activity and immunostaining of ESC markers (SSEA-1, Nanog, and Sox2) of a selected iPS cell line. Scale bar, 200 μm . (C) Activation of endogenous *Oct4*, *Sox2* and *Nanog* in iPS cells derived from reprogramming by OSK plus Kdm2b. RT-qPCR were carried out on RNA samples harvested from MEF, ESC and selected iPS cell (iPSC) lines. (D) Silencing of retroviral transgenes pMXs-*Oct4*, -*Sox2* and -*Klf4* and lentiviral inducible pTYF-TRE-*Kdm2b* in the absence of doxycyclin in selected iPS cell lines. The RNA sample from doxycyclin-treated *Oct4-IRES-GFP/Rosa26-M2rtTA* MEFs introduced with the four transgenes serves as positive control (Control). (E) Representative pictures show the presence of gut-like cavities (endoderm), muscle tissues (mesoderm) and neural rosettes (ectoderm) in teratoma derived from iPS cells reprogrammed by OSK plus Kdm2b. Scale bar, 100 μm . (F) iPS cells generated in the presence of Kdm2b are competent in chimera mice generation. Shown are chimera mice derived from two representative iPS cell lines (iPSC8 and 17).

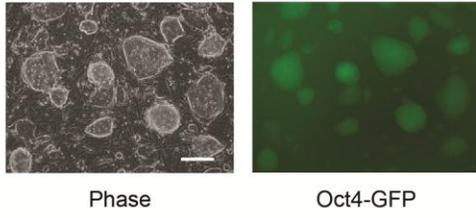
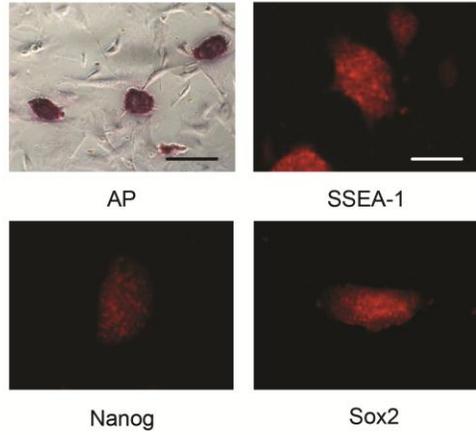
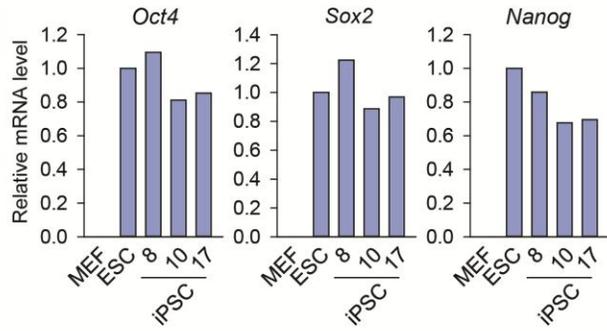
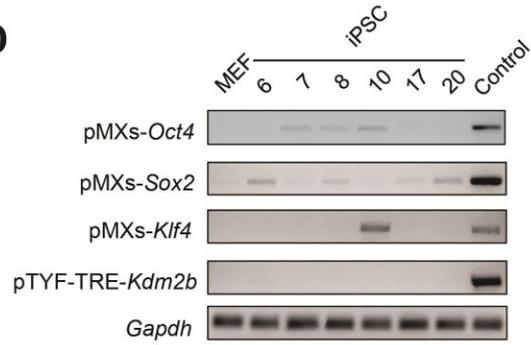
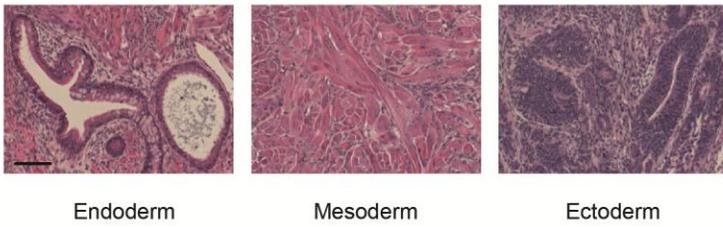
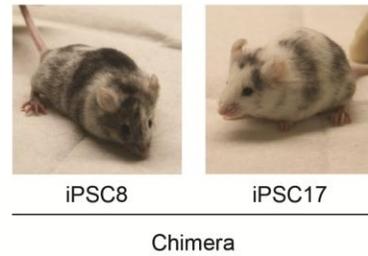
A**B****C****D****E****F**

Figure 3-3 Kdm2b facilitates iPS cell generation in a JmjC and ZF domain-dependent manner.

(A) Western blot analysis demonstrates equal expression of wild-type and mutant Kdm2b with differential effects on global H3K36me2 levels. Lentivirus carrying doxycyclin inducible wild-type and mutant Kdm2b were respectively transduced into MEFs that constitutively expresses the transactivator rtTA. Expression of the constructs was induced by addition of doxycyclin (1 μ g/ml). Western blotting shows the equal expression levels of the various Kdm2b forms as well as their effects on histone modifications. Tubulin and histone H3 serves as a loading control. **(B)** Mutations on JmjC or ZF domain of Kdm2b abrogate its effect on iPS generation. Shown are Oct4-GFP⁺ colony numbers counted at days 12 and 16 reprogrammed using OSK, OSK plus wild-type Kdm2b and its JmjC and ZF mutant. $n = 2$. **(C)** The effect of Kdm2b on proliferation is dependent on its enzymatic activity, but independent of its ZF domain. Cell numbers are monitored for 12 days of reprogramming by OSK, and OSK plus WT and mutant Kdm2b. $n = 2$. **(D)** Relative reprogramming efficiency independent of the cell proliferation effect. The Oct4-GFP⁺ colony numbers in panel B were divided by the total cell numbers in panel c, and the results were all compared to OSK reprogramming and shown as the relative reprogramming efficiency. $n = 2$.

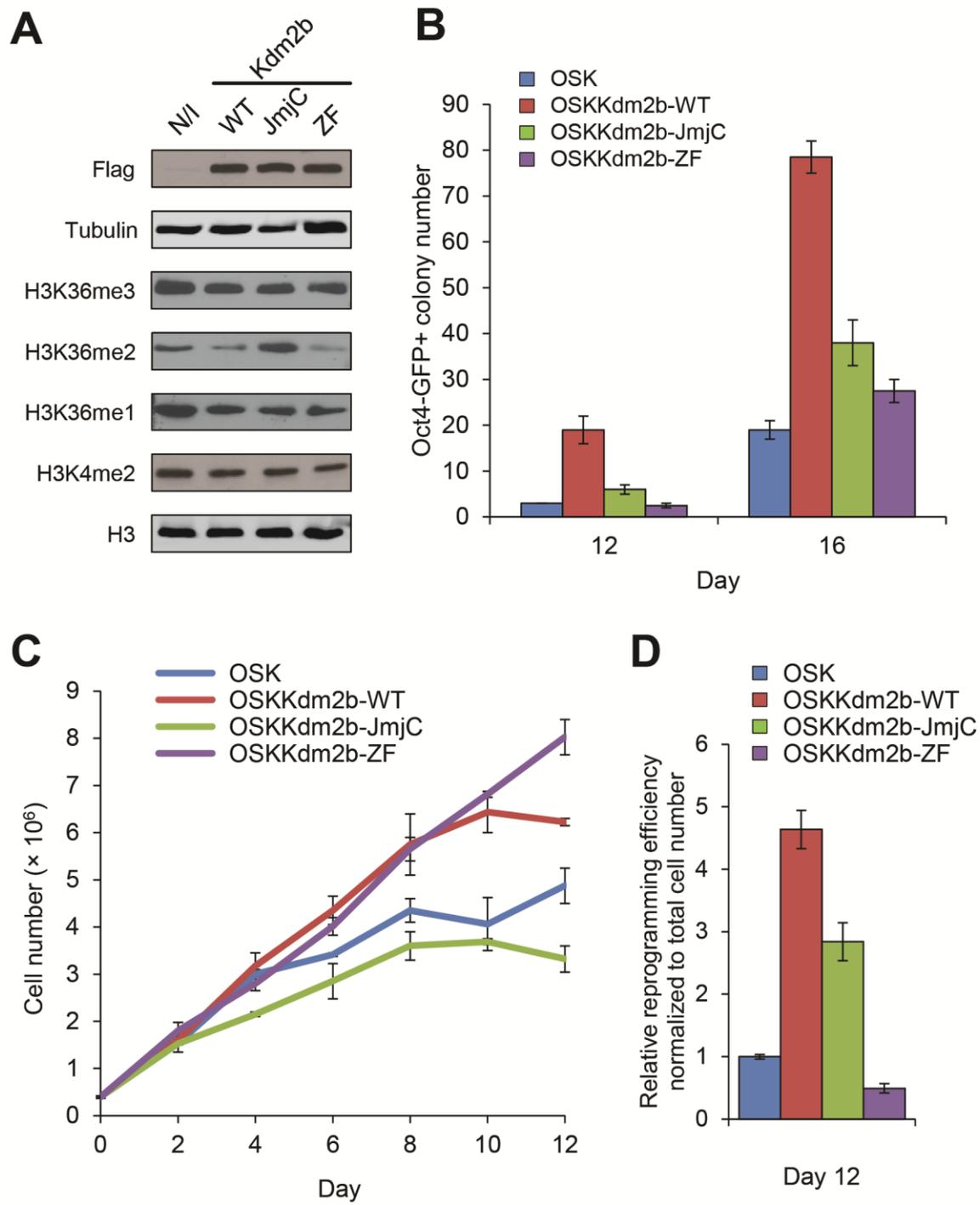


Figure 3-4 Kdm2b promotes iPS cell generation independently of Ink4a/Arf.

(A) RT-qPCR analysis of the expression of Ink4a, Arf and Ink4b in Kdm2b-facilitated reprogramming. RNA samples were harvested from MEF, iPSC and cells at day 4 (D4), 8 (D8) and 12 (D12) during reprogramming by OSK (-) and OSK plus Kdm2b (+). All the qPCR results were normalized to the expression of *Gapdh* and compared to the sample in the absence of Kdm2b (-) at D4. *n* = 2. (B) Western blot analysis shows the protein level of Ink4a and Arf in Kdm2b-facilitated reprogramming at D4 and D8. Protein extracts from the reprogramming cell samples in panel A, as well as MEFs, were used. Tubulin serves as a loading control. (C) RT-qPCR analysis demonstrates depletion of Ink4a and Arf transcript by introduction of shRNA against *Ink4a/Arf*. *n* = 2. (D) Western blot analysis demonstrates efficient depletion of Ink4a and Arf by shRNA against *Ink4a/Arf*. (E) Kdm2b remains its capacity to enhance reprogramming upon Ink4a and Arf depletion. Oct4-GFP+ colony number at day 16 for OSK or OSK plus Kdm2b reprogramming with or without Ink4a/Arf knock-down. *n* = 2.

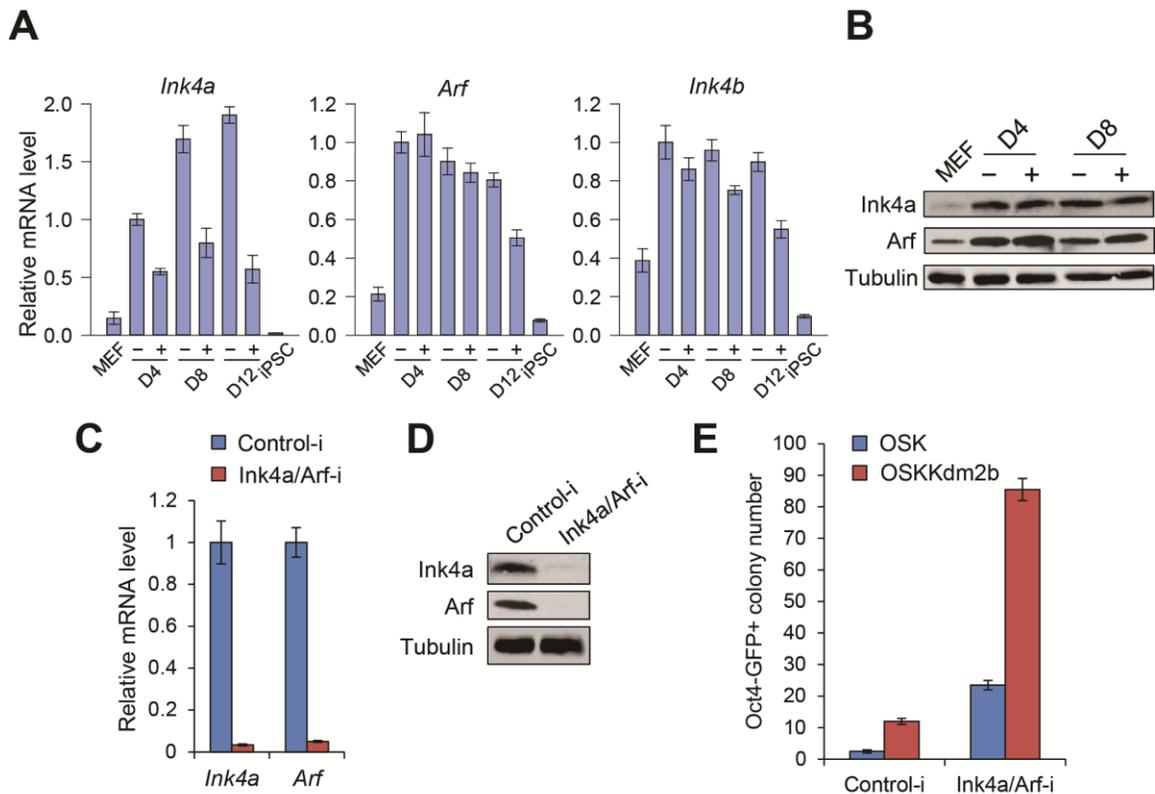


Figure 3-5 Kdm2b exerts its effect from the beginning of reprogramming process and facilitates gene activation in early stage of reprogramming.

(A) Kdm2b functions to promote iPS cell generation from the beginning of reprogramming process. *Oct4-IRES-GFP/Rosa26-rtTAM2* MEFs were transduced with retroviral OSK and lentiviral Kdm2b driven by a doxycyclin inducible promoter. After transduction, doxycyclin (Dox) was applied to (red bars, left panels) or omitted from (grey bars) the cell cultures during the initial 12 days of reprogramming. Oct4-GFP⁺ cells were counted at day 12 and reprogramming efficiencies (green bars, right panels) are compared to that reprogrammed by OSK, which is set to 1. $n = 2$. (B) Hierarchical clustering of genes whose expression are significantly affected by the addition of Kdm2b into the reprogramming cocktail. The expression changes of the selected probes revealed by microarray analysis are at least 2 folds, and the changes occur at least in one of the three time points (days 4, 8 and 12). The heat map was derived via Pearson correlation coefficient from data of two biological replicates. Genes upregulated or downregulated by Kdm2b constitute two distinct clusters, with the upregulated genes subdivided into three groups (I, II and III) based on their time of activation. (C) Gene ontology analysis for all the genes identified in panel B. All the presented terms are above the p -value of 1×10^{-3} . (D) Venn diagram showing the common genes of Kdm2b-affected genes (panel A) and the previously identified signature reprogramming genes. Ten genes, which are represented by 12 probes in the microarray, are shown in the box. (E) Venn diagram showing the distribution of Kdm2b-affected probes at different time of reprogramming. (F) Gene ontology analysis for biological processes for different sections in panel E. The sections without any term enriched above the p -value of 1×10^{-3} are omitted.

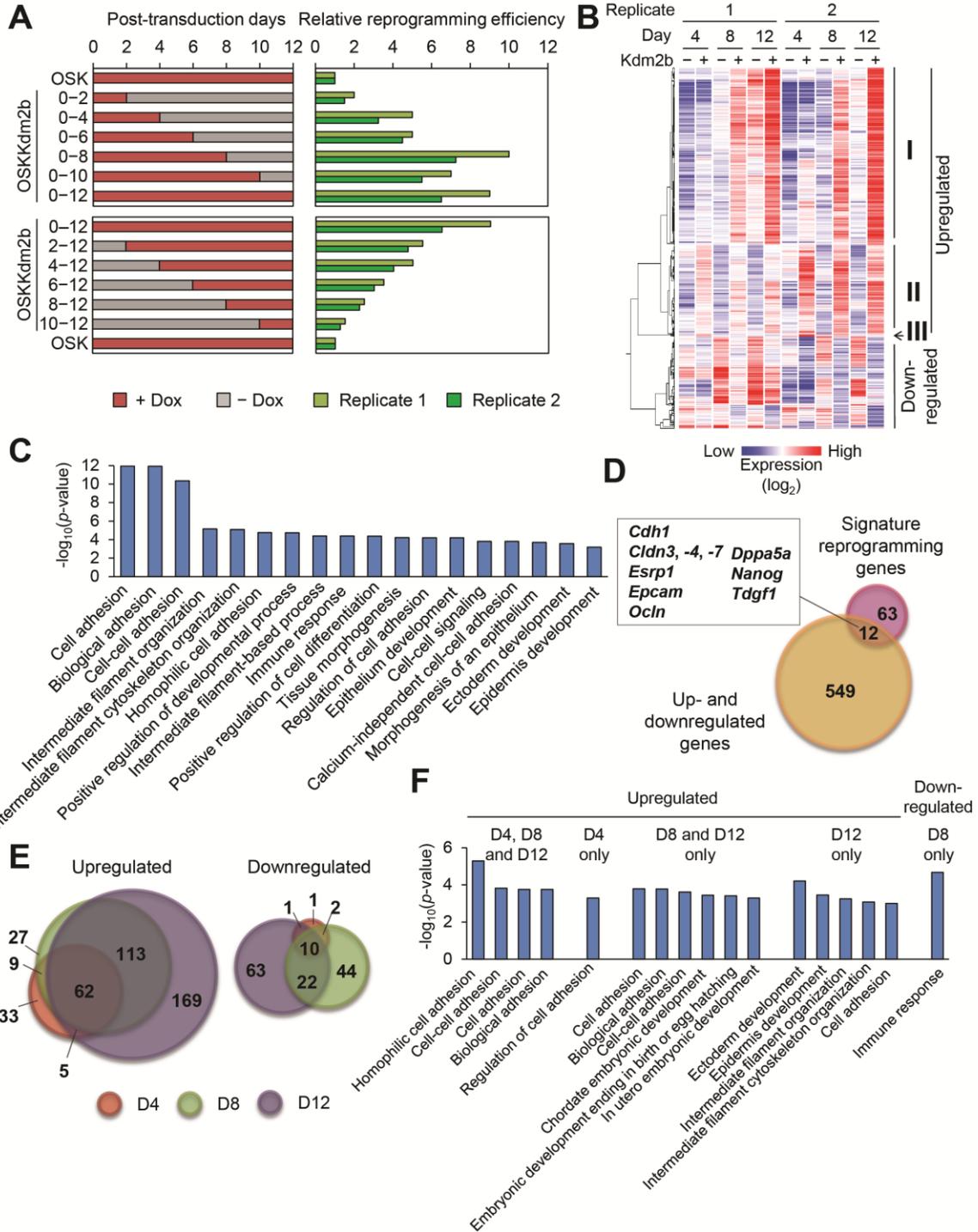


Figure 3-6 Kdm2b directly activates early responsive genes in reprogramming.

(A) RT-qPCR analysis verifying genes upregulated at different time points in Kdm2b-facilitated reprogramming. RNA samples were extracted from MEF, iPSC and cells at day 4 (D4), 8 (D8) and 12 (D12) during reprogramming by OSK (-) and OSK plus Kdm2b (+). The qPCR results were normalized to *Gapdh* and compared to the samples in the absence of Kdm2b (-) at D4. $n = 2$. (B) Epithelial gene expression amplified by Kdm2b at the earliest time point. Expression data from panel A were rescaled to manifest the enhancement by introducing Kdm2b into the OSK cocktail (-) at day 4 (D4). The expression of OSK reprogramming (+) is set to 1. $n = 2$. (C) Western blot analysis demonstrates upregulation of *Cdh1* and *Epcam* by Kdm2b. Protein extracts from the reprogramming cell samples in panel a, as well as MEFs and ESCs, were used. Tubulin serves as a loading control. (D) RT-qPCR analysis demonstrates that Kdm2b-enhanced activation of *Cdh1*, *Epcam* and *Nanog* depends on its JmjC and ZF domains. The analysis is performed at day 12 of reprogramming with cells reprogrammed by OSK, or OSK plus wild-type Kdm2b (WT), its JmjC or ZF mutants. All the qPCR results were normalized to *Gapdh* and compared to the OSK reprogramming. $n = 2$. (E) Western blot analysis of *Cdh1*, *Epcam* and tubulin with protein extract from the reprogramming cell samples in panel C. (F) Kdm2b cooperate with Oct4, Sox2 and/or Klf4 in transcription activation. RT-qPCR analysis of the expression of *Cdh1*, *Crb3*, *Epcam* and *Nanog* in MEF cells transduced with individual factors of Kdm2b, Oct4 (O), Sox2 (S) and Klf4 (K), as well as different factor combinations. RNA samples were harvested at post-transduction day 12. The qPCR results were first normalized to *Gapdh* and then compared to the sample transduced by OSK. $n = 2$.

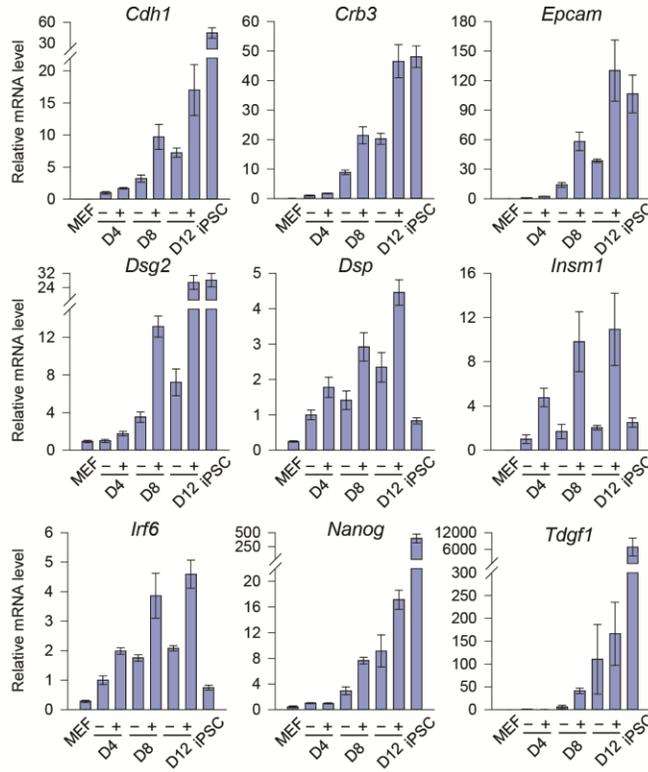
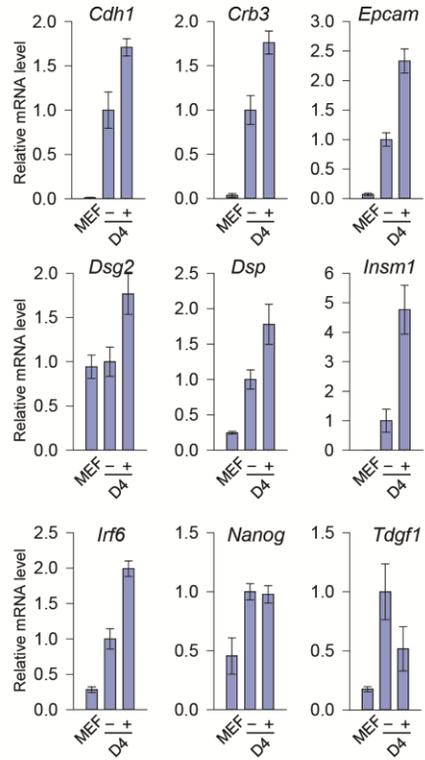
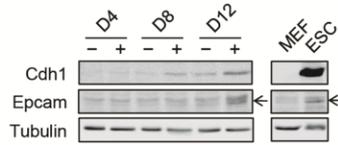
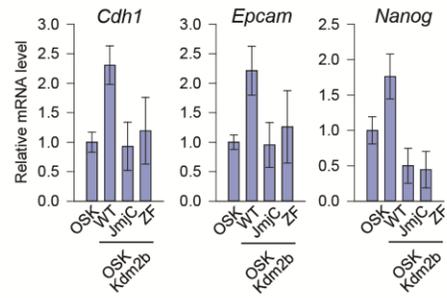
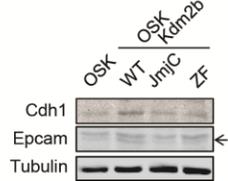
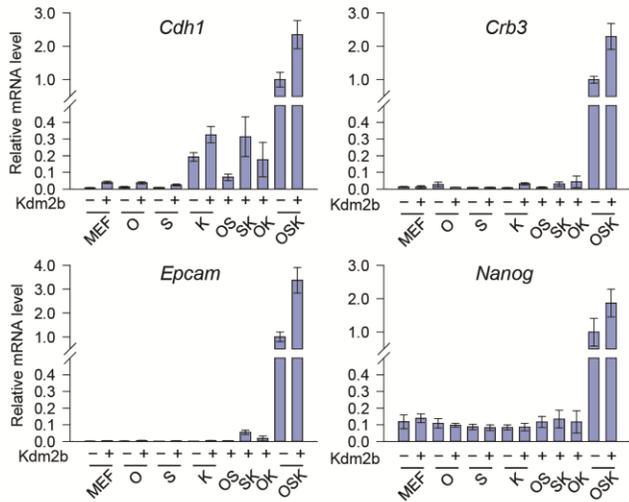
A**B****C****D****E****F**

Figure 3-7 Kdm2b localizes on and demethylates the promoter region of early activated genes.

(A) Chromatin immunoprecipitation (ChIP) analysis demonstrates that Kdm2b directly binds to the promoters of early activated genes. ChIP using anti-Flag antibody and IgG were performed with cells harvested at day 4 of reprogramming using OSK plus Flag-tagged Kdm2b. The qPCR results were compared to IgG controls. $n = 2$. (B) Diagram of *Cdh1* locus illustrating the locations of amplicons relative to the transcription start site (arrow) and transcribed sequence (gray box). (C) Kdm2b localizes from the promoter region to the middle of gene body at the *Cdh1* locus. ChIP was prepared and the results shown as panel A. $n = 2$. (D) ChIP analysis demonstrates reduction of H3K36me2 level correlates with Kdm2b binding to the early activated genes. ChIP using anti-H3K36me2 and H3K4me2 were performed with OSK or OSKKdm2b-introduced cells harvested at day 4. The results for H3K36me2 and H3K4me2 were normalized to IgG. H3K4me2, a modification irrelevant to the enzymatic activity of Kdm2b, serves as a control. $n = 2$.

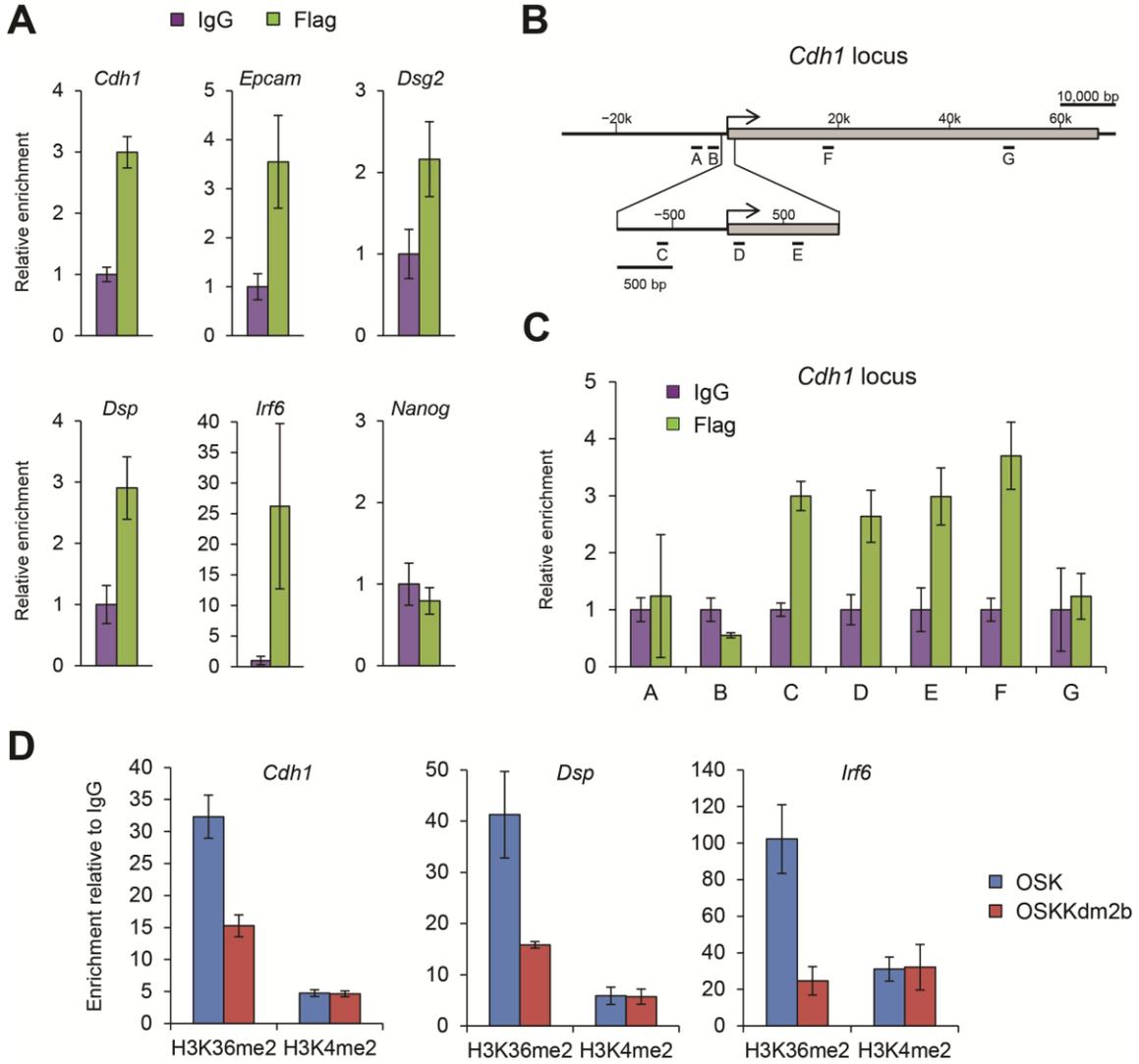


Figure 3-8 Inhibition of the expression of early-activated genes compromises the capacity of Kdm2b to enhance iPS cell generation.

(A) Upregulation of early activated genes in reprogramming by Kdm2b is abrogated by administration of Tgf- β . RNA samples were harvested from reprogramming cells with indicated treatments at day 12. $n = 2$. (B) The capacity of Kdm2b to increase iPS cell generation efficiency is negated by Tgf- β . Oct4-GFP+ colony numbers were calculated at day 16 of reprogramming. $n = 2$. (C) RT-qPCR and (D) Western blotting confirmation of Cdh1 depletion. RNA and protein extracts are harvested from ES cells transduced with shRNA against Cdh1 (Cdh1-i) or control shRNA (Control-i). (E) Inhibition of Cdh1 by shRNA compromises Kdm2b-mediated enhancement of iPS cell generation. Shown are Oct4-GFP+ colony numbers at day 16 of reprogramming. $n = 2$. (F) Transcription cascade model explaining how Kdm2b enhances iPS cell generation. Introduction of Kdm2b into the reprogramming cocktail leads to upregulation of epithelial genes that are activated early during reprogramming, which in turn causes an amplified transcription cascade that enhances the activation of Nanog and the generation of iPS cells.

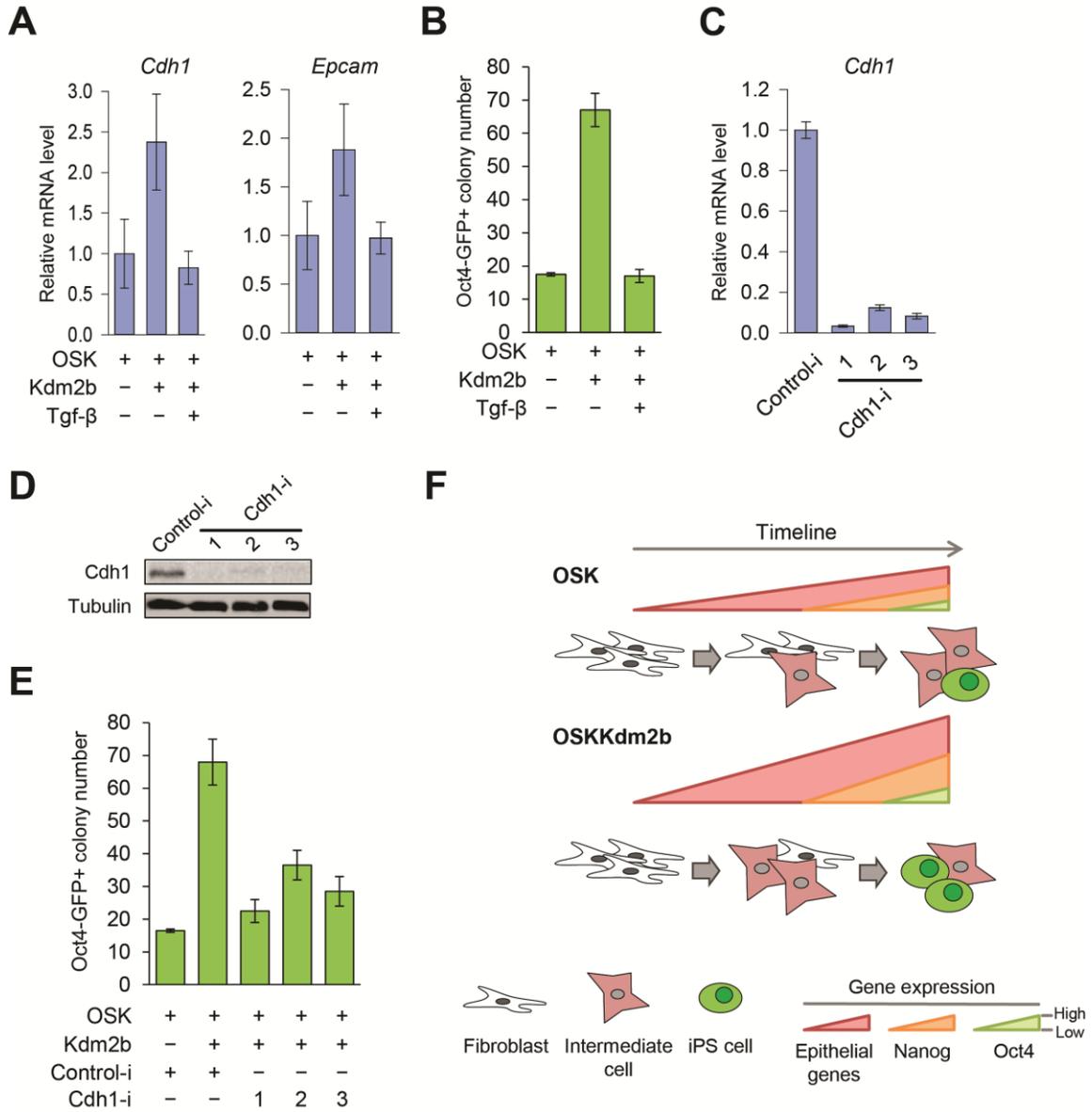


Figure S3-1 Diagram of domain structure for the Kdm2 family members.

The Kdm2 family includes two members, Kdm2a and Kdm2b, with the latter having three isoforms (IF1, IF2 and IF3). The IF1 of Kdm2b is the longest isoform; IF2 lacks the Jumonji-C domain (JmjC) compared to IF1 and IF3; and IF3 is only slightly shorter than IF1 on the N-terminus. The N-terminuses of IF2 and IF3 are distinct from IF1 due to alternative transcription start sites. Both the IF1 and IF3 of Kdm2b share all the functional domains with Kdm2a, including JmjC domain, zinc finger-CXXC (ZF) domain, plant homeo domain (PHD), F-box domain (Fbox) and three runs of leucine-rich repeats (LRR). The number of amino acids (AA) for each protein is indicated in the figure.

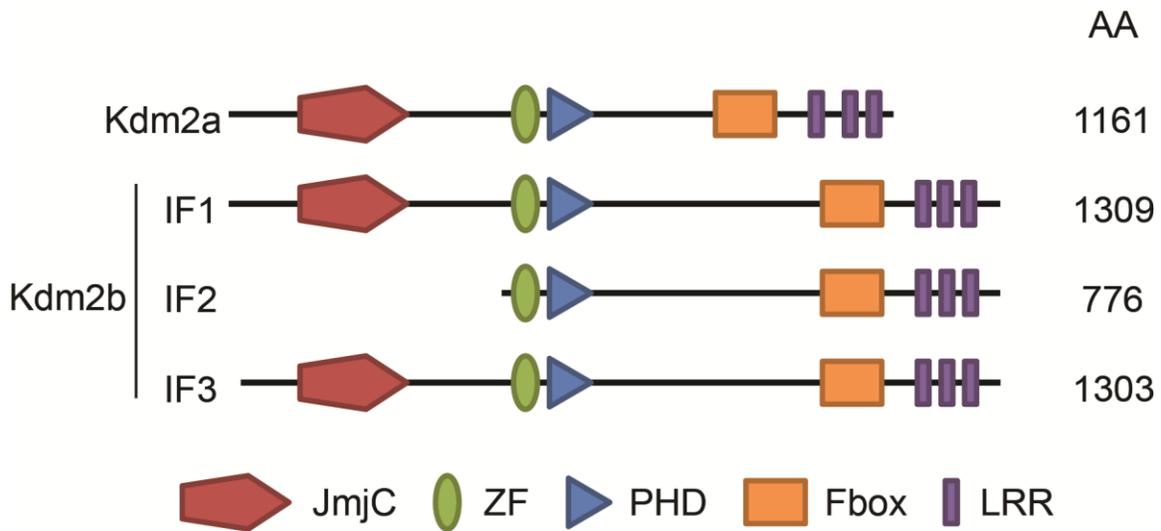


Figure S3-2 Induction of Kdm2b in reprogramming and ectopic expression of Kdm2b.

(A) Induction of endogenous *Kdm2b* isoform 1 (IF1) during OSK reprogramming. Samples from OSK reprogramming at post-transduction day 4 (D4), 8 (D8) and 12 (D12), as well as MEF, ESC and iPSC, were subjected to RT-qPCR specifically detecting endogenous *Kdm2b-IF1*. The results were normalized to the expression of *Gapdh* and compared to MEF. $n = 2$. (B) Doxycyclin-inducible overexpression of Kdm2b introduced by lentivirus. RNA samples were harvested from ESC, MEF without transduction (-) and MEF transduced with lentiviral Kdm2b construct and treated with doxycyclin for 4 days (OE). The results of RT-qPCR detecting the open reading frame (ORF) of *Kdm2b-IF1* were normalized to the expression of *Gapdh* and compared to the expression level in ESC, which is set to 1. $n = 2$.

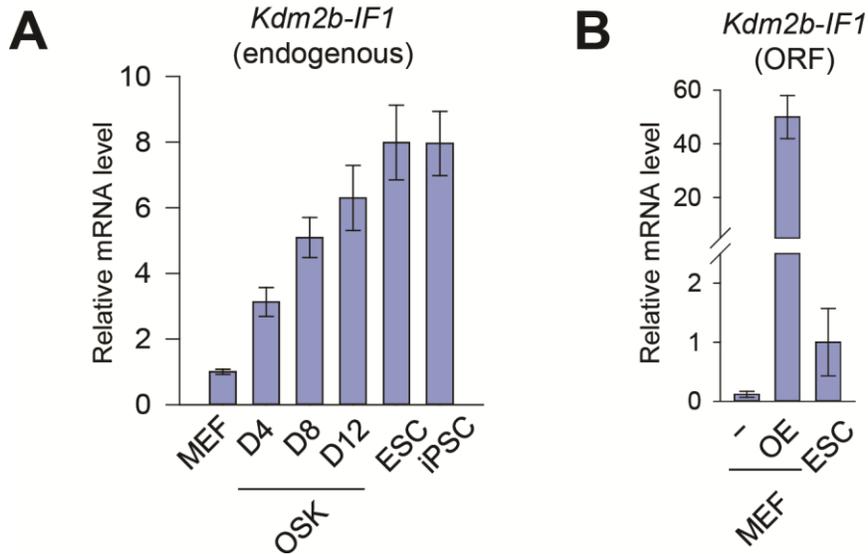


Figure S3-3 Overexpression of Kdm2b does not affect expression of retroviral reprogramming factors.

RNA samples were harvested from MEF introduced with indicated factors. RT-qPCRs were performed using primers specifically detecting the transgenes. The results were normalized to *Gapdh* and compared to OSK or OSKM transduction. $n = 2$.

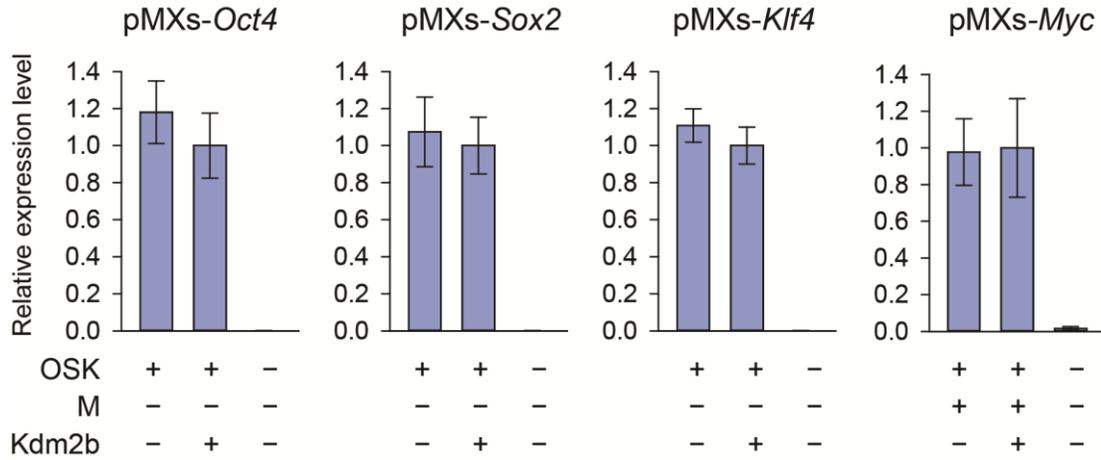


Figure S3-4 Kdm2b enhances transcription activation in the examined time window.

Hierarchical clustering of all the probes representing upregulated and downregulated signature reprogramming genes based on their expression in OSK and Kdm2b-assisted OSK reprogramming. The signature genes were compiled from the previously published literature (Samavarchi-Tehrani et al. 2010) and denoted on the right side of the heat map. Kdm2b mainly affects the expression of upregulated genes in the initial phase in the examined time window from day 4 to day 12. The upregulated genes in the maturation and stabilization phase and the downregulated genes in the whole reprogramming process clustered together with mild or no alternation in gene expression. The extent of upregulation by Kdm2b from low (white) to high (orange) is also indicated by the vertical wedge with color gradient. Overexpression of Kdm2b does not affect expression of retroviral reprogramming factors.

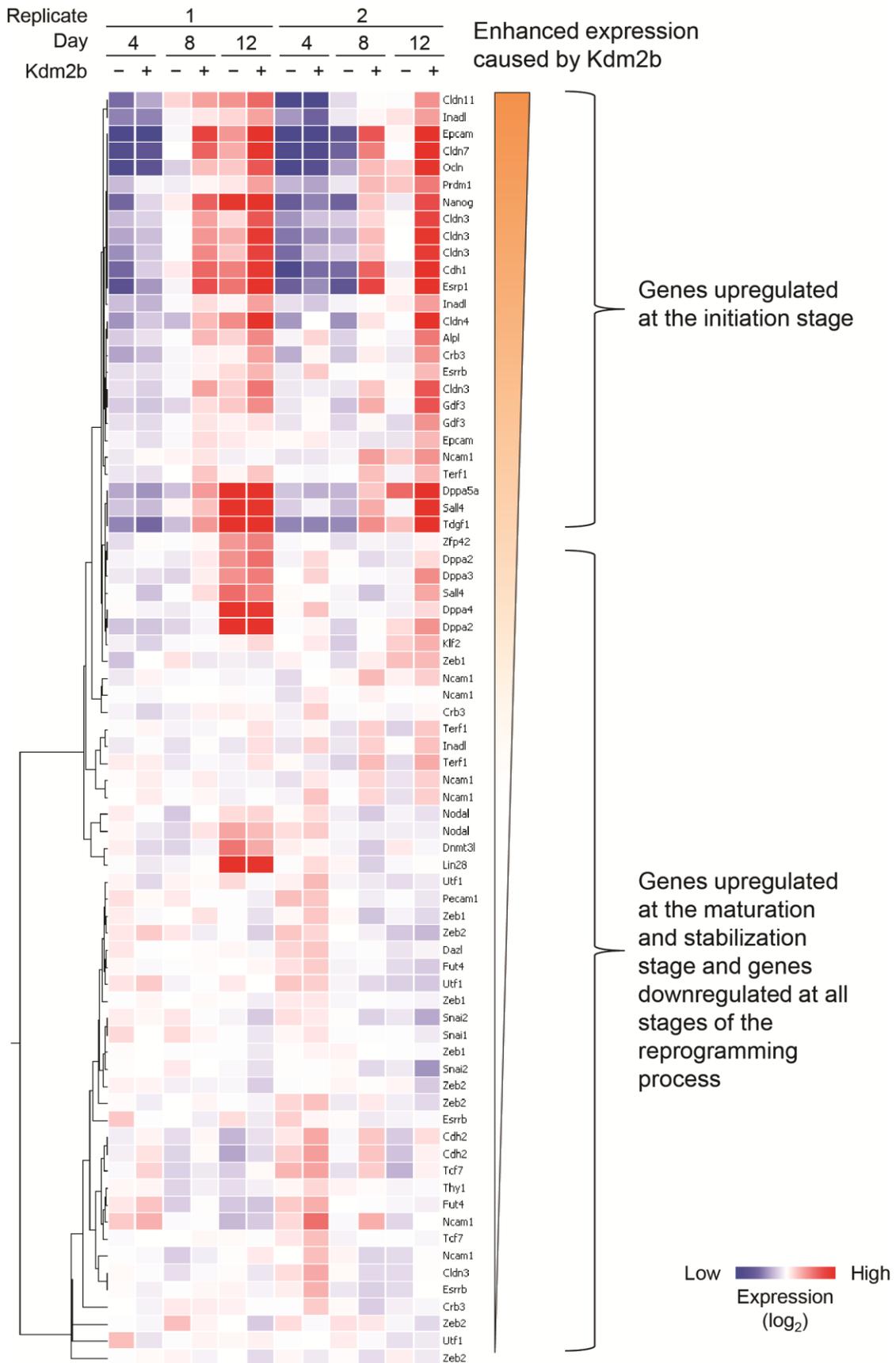
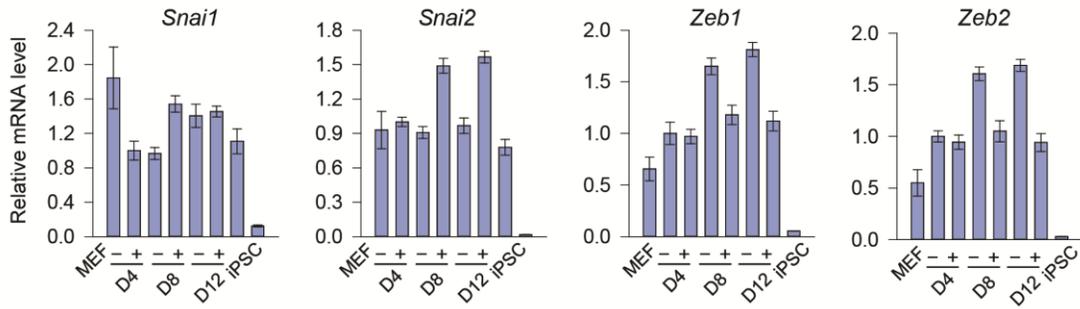


Figure S3-5 RT-qPCR showing the expression of mesenchymal genes and endogenous pluripotent genes in Kdm2b-assisted reprogramming.

(A) Mesenchymal genes repressed in the maturation stage in reprogramming show mild to slight repression at the later time points in the Kdm2b-assisted reprogramming. (B) Endogenous loci of *Oct4* and *Sox2* do not show activation by Kdm2b in the examined time points. RNA samples for RT-qPCR in A and B were prepared as described in Figure 3-6A. $n = 2$.

A



B

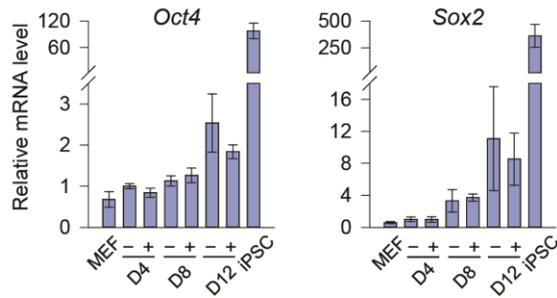


Table S3-1 Target sequences of shRNAs used in Chapter Three.

Target	Seed sequence	Reference
Kdm2b	GCTCCA ACTCAGTTACTGT	(He et al. 2008)
Control (<i>lacZ</i> from <i>E. coli</i>)	G TTCAGATGTGCGGCGAGT	
Ink4a/Arf	GCGATATTTGCGTTCCGCT	This study
Cdh1 (i1)	GGAGATGCAGAATAATTAT	(Li et al. 2010)
Cdh1 (i2)	GCTGGAATCTTTGTCCATGTA	
Cdh1 (i3)	GCAGGAAGAGAACATTCTA	

Table S3-2 Sequences of primers used in quantitative and semi-quantitative RT-PCR in Chapter Three.

Quantitative RT-PCR		
Target	Direction	Sequence
<i>Arf</i>	F	GCCGCACCGGAATCCT
	R	TTGAGCAGAAGAGCTGCTACGT
<i>Cdh1</i>	F	AACAACCTGCATGAAGGCGGGAATC
	R	CCTGTGCAGCTGGCTCAAATCAA
<i>Crb3</i>	F	CCGGACCCTTTCACAAATAG
	R	TCGCATGAGCAGAAACAGTC
<i>Dsp</i>	F	GCGGCTTGGGTGGGAGTGTC
	R	GGCCGCAATGGGGCTTGACT
<i>Dsg</i>	F	AAGGCCACGCCATTCCCATCA
	R	TTTCCAATCGACCTGCTGAGGCT
<i>Gapdh</i>	F	CATGGCCTTCCGTGTTCTA
	R	GCCTGCTTCACCACCTTCTT
<i>Ink4a</i>	F	CGTACCCCGATTTCAGGTGAT
	R	TTGAGCAGAAGAGCTGCTACGT
<i>Ink4b</i>	F	ATGTTGGGCGGCAGCAGTGACG
	R	ATCTCCAGTGGCAGCGTGACG
<i>Insm1</i>	F	TGCGTCCGGCCTGCTAGAGT
	R	GCTCCACCGAAGCGAAGCGA
<i>Irf6</i>	F	GCACCTGCTCCTGAGCACGG
	R	GCCGCGTGGCATGTTTCCAG
<i>Kdm2a</i>	F	AGGAAGGCATCCCTGGAGTGGTT
	R	AGCGACGACGCATGGTACCACG

<i>Kdm2b-IF1</i> (en)	F	ATGCAGGGGGCAGGAGTGTTGA
	R	CTGAAGCCACGGACGCTGACAA
<i>Kdm2b-IF1</i> (ORF)	F	TGGAGGCAGAGAAAGACTCTGGAAG
	R	CCAGGCTGAAGCCACGGACG
<i>Kdm2b-IF2</i>	F	GGCGCTGCTCCAGCCATGTCAG
	R	CGAGCTCCTGCTGTTGTTCGGTT
<i>Kdm2b-IF3</i>	F	CGCTCGTGGGGAGATGAGGCAG
	R	GCCACGGACGCTGACAATTCCT
<i>Nanog</i>	F	AAGCAGAAGATGCGGACTGT
	R	ATCTGCTGGAGGCTGAGGTA
<i>Ocln</i>	F	CCTCCAATGGCAAAGTGAATGGCA
	R	TGTTTCATAGTGGTCAGGGTCCGT
<i>Oct4</i> (en)	F	TAGGTGAGCCGTCTTTCCAC
	R	CCTTGGAAGCTTAGCCAGGT
pMXs- <i>Klf4</i>	F	TCCCAGTGTGGTGGTACGGGA
	R	AGTTGCTTTCCACTCGTGCT
pMXs- <i>Myc</i>	F	GCCACCGCCTACATCCTGTCC
	R	GACCGGCGCTCAGCTGGAAT
pMXs- <i>Oct4</i>	F	(As pMXs- <i>Klf4</i> F)
	R	AGTTGCTTTCCACTCGTGCT
pMXs- <i>Sox2</i>	F	CTGCCCTGTGCGACATGTG
	R	CTTTTATTTTATCGTCGACC
<i>Snai1</i>	F	TTGTGTCTGCACGACCTGTGGAAA
	R	TCTTCACATCCGAGTGGGTTTGGGA
<i>Snai2</i>	F	CACATTCGAACCCACACATTGCCT
	R	TGTGCCCTCAGGTTTGATCTGTCT

<i>Sox2</i> (en)	F	TGATCAGCATGTACCTCCCC
	R	CCCTCCCAATTCCCTTGTAT
<i>Tdgf1</i>	F	CGGAGATCTTGGCTGCTAAC
	R	CTTCGACGGCTCGTAAAAAC
<i>Zeb1</i>	F	TGCTCACCTGCCCCGTATTGTGATA
	R	AGTGCACCTTGAACCTGCGGTTTCC
<i>Zeb2</i>	F	TGATAGCCTTGCAAACCCTCTGGA
	R	ATTGTGGTCTGGATCGTGGCTTCT

Semi-quantitative RT-PCR

Target	Direction	Sequence
pMXs- <i>Klf4</i>	F	CCAAAGAGGGGAAGAAGGTC
	R	CCTACAGGTGGGGTCTTTCA
pMXs- <i>Oct4</i>	F	CCAGAAGGGCAAAAGATCAA
	R	(As pMXs- <i>Klf4</i> R)
pMXs- <i>Sox2</i>	F	ACCAGCTCGCAGACCTACAT
	R	(As pMXs- <i>Klf4</i> R)
pTYF-TRE- <i>Kdm2b</i>	F	CCCGAGACTCGCTGACAG
	R	GGCTAAGATCTACAGCTGCC
<i>Gapdh</i>	F	CATGGCCTTCCGTGTTCCCTA
	R	GCCTGCTTACCACCTTCTT

F, forward primers; R, reverse primers; en, endogenous loci; ORF, open reading frame

Table S3-3 Sequences of primers used in ChIP-qPCR in Chapter Three.

Target locus	Amplicon location	Direction	Sequence
<i>Cdh1</i>	A: -5,504 ~ -5,383	F	GGGGCTACCAACCGCAAGGTG
		R	AGGCCGGCCCCTTCCCATAG
	B: -1,935 ~ -1,770	F	AGGCACAGCCTCCACACCTCA
		R	AGCCACCCTCCTCAGCGACA
	C: -646 ~ -563	F	CCCCACTTGTGCAATCCCAGCA
		R	CCAACTCAGGTGGGCCTGGA
	D: -10 ~ 66	F	TGCGCTGCTCACTGGTGTGG
		R	AGGCCGGGCAGGAGTCTAGC
	E: 415 ~ 505	F	CCGGGTTGAGCAGGGTCCCT
		R	CCGCGCTACTTCAGCACCCC
	F: 16,297 ~ 16,377	F	AGCTGTTCGGAGCCTCAGGGG
		R	TGTGGCCCTCCCCTCTCAGC
	G: 50,234 ~ 50,315	F	TCAGCTGCCCCGAAAATGA
		R	GGGCAGAGCCTGCCACCAAC
<i>Dsg2</i>	-287 ~ -188	F	CGCTTGGCCAGCGTGTCTCC
		R	GCGCGACTCATGGAGCCTGG
<i>Dsp</i>	-251 ~ -172	F	TCCGCCTCTTGGGGGAGTAGG
		R	GACCCCGCGGGACTTGCTTC
<i>Epcam</i>	-217 ~ -70	F	GGGCCTGCTTTTTCTCCCGCC
		R	GCCCCAGTTTGGGTTGGTCAG
<i>Irf6</i>	-383 ~ -226	F	TAGAACACGGGGTCCACGGGC
		R	GAGCGTCGCCCTCCCGTA
<i>Nanog</i>	-180 ~ -134	F	TCCCTCCCTCCCAGTCTG
		R	CCTCCTACCCTACCCACCC

REFERENCE

- Ambrosetti DC, Scholer HR, Dailey L, Basilico C. 2000. Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. *J Biol Chem* **275**: 23387-23397.
- Ang YS, Tsai SY, Lee DF, Monk J, Su J, Ratnakumar K, Ding J, Ge Y, Darr H, Chang B et al. 2011. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**: 183-197.
- 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**: 126-140.
- Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M et al. 2006. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8**: 532-538.
- Banaszynski LA, Allis CD, Lewis PW. 2010. Histone variants in metazoan development. *Dev Cell* **19**: 662-674.
- Banito A, Rashid ST, Acosta JC, Li S, Pereira CF, Geti I, Pinho S, Silva JC, Azuara V, Walsh M et al. 2009. Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev* **23**: 2134-2139.
- Bernstein BE, Meissner A, Lander ES. 2007. The mammalian epigenome. *Cell* **128**: 669-681.
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315-326.
- Bernstein E, Allis CD. 2005. RNA meets chromatin. *Genes Dev* **19**: 1635-1655.
- Bilodeau S, Kagey MH, Frampton GM, Rahl PB, Young RA. 2009. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev* **23**: 2484-2489.
- Black BE, Bassett EA. 2008. The histone variant CENP-A and centromere specification. *Curr Opin Cell Biol* **20**: 91-100.

- Blackledge NP, Klose R. 2011. CpG island chromatin: a platform for gene regulation. *Epigenetics* **6**: 147-152.
- 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**: 179-190.
- Boland MJ, Hazen JL, Nazor KL, Rodriguez AR, Gifford W, Martin G, Kupriyanov S, Baldwin KK. 2009. Adult mice generated from induced pluripotent stem cells. *Nature* **461**: 91-94.
- Bolden JE, Peart MJ, Johnstone RW. 2006. Anticancer activities of histone deacetylase inhibitors. *Nat Rev Drug Discov* **5**: 769-784.
- Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**: 947-956.
- Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK et al. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**: 349-353.
- Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K. 2006. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* **20**: 1123-1136.
- Brambrink T, Foreman R, Welstead GG, Lengner CJ, Wernig M, Suh H, Jaenisch R. 2008. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2**: 151-159.
- Cairns BR. 2009. The logic of chromatin architecture and remodelling at promoters. *Nature* **461**: 193-198.
- Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**: 1039-1043.
- Chamberlain SJ, Yee D, Magnuson T. 2008. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells* **26**: 1496-1505.
- 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**: 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**: 1230-1234.

- Chen T, Yuan D, Wei B, Jiang J, Kang J, Ling K, Gu Y, Li J, Xiao L, Pei G. 2010. E-cadherin-mediated cell-cell contact is critical for induced pluripotent stem cell generation. *Stem Cells* **28**: 1315-1325.
- Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**: 1106-1117.
- Cole MF, Johnstone SE, Newman JJ, Kagey MH, Young RA. 2008. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev* **22**: 746-755.
- Davis RL, Weintraub H, Lassar AB. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**: 987-1000.
- Deaton AM, Bird A. 2011. CpG islands and the regulation of transcription. *Genes Dev* **25**: 1010-1022.
- Edwards CA, Ferguson-Smith AC. 2007. Mechanisms regulating imprinted genes in clusters. *Curr Opin Cell Biol* **19**: 281-289.
- Eminli S, Foudi A, Stadtfeld M, Maherali N, Ahfeldt T, Mostoslavsky G, Hock H, Hochedlinger K. 2009. Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nat Genet* **41**: 968-976.
- Esteban MA, Wang T, Qin B, Yang J, Qin D, Cai J, Li W, Weng Z, Chen J, Ni S et al. 2010. Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* **6**: 71-79.
- Farkas LM, Haffner C, Giger T, Khaitovich P, Nowick K, Birchmeier C, Paabo S, Huttner WB. 2008. Insulinoma-associated 1 has a panneurogenic role and promotes the generation and expansion of basal progenitors in the developing mouse neocortex. *Neuron* **60**: 40-55.
- Fazio TG, Huff JT, Panning B. 2008. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* **134**: 162-174.
- Feng B, Ng JH, Heng JC, Ng HH. 2009. Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. *Cell Stem Cell* **4**: 301-312.
- Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**: 398-402.

- Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S. 2003. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* **17**: 1870-1881.
- Fouse SD, Shen Y, Pellegrini M, Cole S, Meissner A, Van Neste L, Jaenisch R, Fan G. 2008. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**: 160-169.
- Gamble MJ, Kraus WL. 2010. Multiple facets of the unique histone variant macroH2A: from genomics to cell biology. *Cell Cycle* **9**: 2568-2574.
- Gaspar-Maia A, Alajem A, Polesso F, Sridharan R, Mason MJ, Heidersbach A, Ramalho-Santos J, McManus MT, Plath K, Meshorer E et al. 2009. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* **460**: 863-868.
- Gierl MS, Karoulias N, Wende H, Strehle M, Birchmeier C. 2006. The zinc-finger factor Insm1 (IA-1) is essential for the development of pancreatic beta cells and intestinal endocrine cells. *Genes Dev* **20**: 2465-2478.
- Goldberg AD, Allis CD, Bernstein E. 2007. Epigenetics: a landscape takes shape. *Cell* **128**: 635-638.
- Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* **74**: 481-514.
- 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**: 77-88.
- Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creighton MP, van Oudenaarden A, Jaenisch R. 2009. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**: 595-601.
- Hanna JH, Saha K, Jaenisch R. 2010. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* **143**: 508-525.
- Hargreaves DC, Horng T, Medzhitov R. 2009. Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* **138**: 129-145.
- He J, Kallin EM, Tsukada Y, Zhang Y. 2008. The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b). *Nat Struct Mol Biol* **15**: 1169-1175.
- Heard E, Disteche CM. 2006. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev* **20**: 1848-1867.

- Heldin CH, Landstrom M, Moustakas A. 2009. Mechanism of TGF-beta signaling to growth arrest, apoptosis, and epithelial-mesenchymal transition. *Curr Opin Cell Biol* **21**: 166-176.
- Henikoff S, Shilatifard A. 2011. Histone modification: cause or cog? *Trends Genet* **27**: 389-396.
- Ho L, Crabtree GR. 2010. Chromatin remodelling during development. *Nature* **463**: 474-484.
- Ho L, Ronan JL, Wu J, Staahl BT, Chen L, Kuo A, Lessard J, Nesvizhskii AI, Ranish J, Crabtree GR. 2009. An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A* **106**: 5181-5186.
- Hong H, Takahashi K, Ichisaka T, Aoi T, Kanagawa O, Nakagawa M, Okita K, Yamanaka S. 2009. Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* **460**: 1132-1135.
- Huangfu D, Maehr R, Guo W, Eijkelenboom A, Snitow M, Chen AE, Melton DA. 2008a. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat Biotechnol* **26**: 795-797.
- Huangfu D, Osafune K, Maehr R, Guo W, Eijkelenboom A, Chen S, Muhlestein W, Melton DA. 2008b. Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nat Biotechnol* **26**: 1269-1275.
- Ichida JK, Blanchard J, Lam K, Son EY, Chung JE, Egli D, Loh KM, Carter AC, Di Giorgio FP, Koszka K et al. 2009. A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. *Cell Stem Cell* **5**: 491-503.
- Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP. 2010. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* **6**.
- Jackson-Grusby L, Beard C, Possemato R, Tudor M, Fambrough D, Csankovszki G, Dausman J, Lee P, Wilson C, Lander E et al. 2001. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat Genet* **27**: 31-39.
- Jackson M, Krassowska A, Gilbert N, Chevassut T, Forrester L, Ansell J, Ramsahoye B. 2004. Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Mol Cell Biol* **24**: 8862-8871.
- Jaenisch R, Young R. 2008. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**: 567-582.

- Jiang H, Shukla A, Wang X, Chen WY, Bernstein BE, Roeder RG. 2011. Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell* **144**: 513-525.
- 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**: 353-360.
- Kang L, Wang J, Zhang Y, Kou Z, Gao S. 2009. iPS cells can support full-term development of tetraploid blastocyst-complemented embryos. *Cell Stem Cell* **5**: 135-138.
- Kawamura T, Suzuki J, Wang YV, Menendez S, Morera LB, Raya A, Wahl GM, Belmonte JC. 2009. Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* **460**: 1140-1144.
- Kim J, Chu J, Shen X, Wang J, Orkin SH. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**: 1049-1061.
- Klose RJ, Bird AP. 2006. Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* **31**: 89-97.
- Knoepfler PS. 2008. Why myc? An unexpected ingredient in the stem cell cocktail. *Cell Stem Cell* **2**: 18-21.
- Kornberg RD, Lorch Y. 1999. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**: 285-294.
- Kouzarides T. 2007. Chromatin modifications and their function. *Cell* **128**: 693-705.
- Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS et al. 2008. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**: e1000242.
- Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**: 116-120.
- Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K et al. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**: 301-313.
- Lessard J, Wu JI, Ranish JA, Wan M, Winslow MM, Staahl BT, Wu H, Aebersold R, Graef IA, Crabtree GR. 2007. An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55**: 201-215.
- Li E, Bestor TH, Jaenisch R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**: 915-926.

- 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**: 1136-1139.
- Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, Patel DJ. 2006. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**: 91-95.
- Li R, Liang J, Ni S, Zhou T, Qing X, Li H, He W, Chen J, Li F, Zhuang Q et al. 2010. A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7**: 51-63.
- Liang G, Taranova O, Xia K, Zhang Y. 2010. Butyrate promotes induced pluripotent stem cell generation. *J Biol Chem* **285**: 25516-25521.
- Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J et al. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**: 431-440.
- Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**: 251-260.
- Maherali N, Hochedlinger K. 2009. Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. *Curr Biol* **19**: 1718-1723.
- Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, Stadtfeld M, Yachechko R, Tchieu J, Jaenisch R et al. 2007a. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**: 55-70.
- . 2007b. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cells* **1**: 55-70.
- Mali P, Chou BK, Yen J, Ye Z, Zou J, Dowey S, Brodsky RA, Ohm JE, Yu W, Baylín SB et al. 2010. Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. *Stem Cells* **28**: 713-720.
- 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**: 1149-1153.
- Marques M, Laflamme L, Gervais AL, Gaudreau L. 2010. Reconciling the positive and negative roles of histone H2A.Z in gene transcription. *Epigenetics* **5**: 267-272.
- Martin C, Zhang Y. 2007. Mechanisms of epigenetic inheritance. *Curr Opin Cell Biol* **19**: 266-272.

- Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA et al. 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9**: 625-635.
- Meissner A. 2010. Epigenetic modifications in pluripotent and differentiated cells. *Nat Biotechnol* **28**: 1079-1088.
- Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB et al. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**: 766-770.
- Mellitzer G, Bonne S, Luco RF, Van De Casteele M, Lenne-Samuel N, Collombat P, Mansouri A, Lee J, Lan M, Pipeleers D et al. 2006. IA1 is NGN3-dependent and essential for differentiation of the endocrine pancreas. *EMBO J* **25**: 1344-1352.
- Mercer TR, Dinger ME, Mattick JS. 2009. Long non-coding RNAs: insights into functions. *Nat Rev Genet* **10**: 155-159.
- Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, Misteli T. 2006. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Developmental Cell* **10**: 105.
- Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A. 2008. Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**: 49-55.
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**: 553-560.
- Misteli T. 2007. Beyond the sequence: cellular organization of genome function. *Cell* **128**: 787-800.
- 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**: 631-642.
- Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schubeler D. 2008. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* **30**: 755-766.
- Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochizuki Y, Takizawa N, Yamanaka S. 2008. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* **26**: 101-106.

- Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**: 110-113.
- Newmark HL, Lupton JR, Young CW. 1994. Butyrate as a differentiating agent: pharmacokinetics, analogues and current status. *Cancer Lett* **78**: 1-5.
- 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**: 379-391.
- 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**: 372-376.
- O'Carroll D, Erhardt S, Pagani M, Barton SC, Surani MA, Jenuwein T. 2001. The polycomb-group gene *Ezh2* is required for early mouse development. *Mol Cell Biol* **21**: 4330-4336.
- Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**: 247-257.
- Okita K, Ichisaka T, Yamanaka S. 2007. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**: 313-317.
- Orkin SH, Hochedlinger K. 2011. Chromatin connections to pluripotency and cellular reprogramming. *Cell* **145**: 835-850.
- Pasini D, Bracken AP, Jensen MR, Lazzerini Denchi E, Helin K. 2004. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J* **23**: 4061-4071.
- Payer B, Lee JT. 2008. X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* **42**: 733-772.
- Pfau R, Tzatsos A, Kampranis SC, Serebrennikova OB, Bear SE, Tsihchlis PN. 2008. Members of a family of JmjC domain-containing oncoproteins immortalize embryonic fibroblasts via a JmjC domain-dependent process. *Proc Natl Acad Sci U S A* **105**: 1907-1912.
- Plath K, Lowry WE. 2011. Progress in understanding reprogramming to the induced pluripotent state. *Nat Rev Genet* **12**: 253-265.
- Polo JM, Liu S, Figueroa ME, Kulal W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T et al. 2010. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol* **28**: 848-855.

- 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**: 432-445.
- Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, Doty KR, Black JC, Hoffmann A, Carey M, Smale ST. 2009. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**: 114-128.
- Redmer T, Diecke S, Grigoryan T, Quiroga-Negreira A, Birchmeier W, Besser D. 2011. E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. *EMBO Rep* **12**: 720-726.
- Richardson RJ, Dixon J, Jiang R, Dixon MJ. 2009. Integration of IRF6 and Jagged2 signalling is essential for controlling palatal adhesion and fusion competence. *Hum Mol Genet* **18**: 2632-2642.
- Saha A, Wittmeyer J, Cairns BR. 2006. Chromatin remodelling: the industrial revolution of DNA around histones. *Nat Rev Mol Cell Biol* **7**: 437-447.
- Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A, Woltjen K, Nagy A, Wrana JL. 2010. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**: 64-77.
- Schnetz MP, Handoko L, Akhtar-Zaidi B, Bartels CF, Pereira CF, Fisher AG, Adams DJ, Flicek P, Crawford GE, Laframboise T et al. 2010. CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genet* **6**: e1001023.
- Shi Y, Desponts C, Do JT, Hahm HS, Scholer HR, Ding S. 2008a. Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem Cell* **3**: 568-574.
- Shi Y, Do JT, Desponts C, Hahm HS, Scholer HR, Ding S. 2008b. A combined chemical and genetic approach for the generation of induced pluripotent stem cells. *Cell Stem Cell* **2**: 525-528.
- Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**: 844-847.
- Shumacher A, Faust C, Magnuson T. 1996. Positional cloning of a global regulator of anterior-posterior patterning in mice. *Nature* **383**: 250-253.
- 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**: 722-737.

- 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**: 943-955.
- Smith ZD, Nachman I, Regev A, Meissner A. 2010. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nat Biotechnol* **28**: 521-526.
- Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Plath K. 2009a. Role of the Murine Reprogramming Factors in the Induction of Pluripotency. *Cell* **136**: 364-377.
- Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Plath K. 2009b. Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**: 364-377.
- Stadtfield M, Hochedlinger K. 2010. Induced pluripotency: history, mechanisms, and applications. *Genes Dev* **24**: 2239-2263.
- Stadtfield M, Maherali N, Breault DT, Hochedlinger K. 2008. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2**: 230-240.
- Strahl BD, Allis CD. 2000. The language of covalent histone modifications. *Nature* **403**: 41-45.
- Suzuki MM, Bird A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**: 465-476.
- Szenker E, Ray-Gallet D, Almouzni G. 2011. The double face of the histone variant H3.3. *Cell Res* **21**: 421-434.
- Tada M, Tada T, Lefebvre L, Barton SC, Surani MA. 1997. Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *EMBO J* **16**: 6510-6520.
- Tada M, Takahama Y, Abe K, Nakatsuji N, Tada T. 2001. Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Curr Biol* **11**: 1553-1558.
- Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**: 663-676.
- Talbert PB, Henikoff S. 2010. Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* **11**: 264-275.
- Tam WL, Lim CY, Han J, Zhang J, Ang YS, Ng HH, Yang H, Lim B. 2008. T-cell factor 3 regulates embryonic stem cell pluripotency and self-renewal by the transcriptional control of multiple lineage pathways. *Stem Cells* **26**: 2019-2031.

- Thomason HA, Zhou H, Kouwenhoven EN, Dotto GP, Restivo G, Nguyen BC, Little H, Dixon MJ, van Bokhoven H, Dixon J. 2010. Cooperation between the transcription factors p63 and IRF6 is essential to prevent cleft palate in mice. *J Clin Invest* **120**: 1561-1569.
- Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr AR, Deaton A, Andrews R, James KD et al. 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**: 1082-1086.
- Tremethick DJ. 2007. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* **128**: 651-654.
- 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**: 811-816.
- Tzatsos A, Pfau R, Kampranis SC, Tsiichlis PN. 2009. Ndy1/KDM2B immortalizes mouse embryonic fibroblasts by repressing the Ink4a/Arf locus. *Proc Natl Acad Sci U S A* **106**: 2641-2646.
- Utikal J, Polo JM, Stadtfeld M, Maherali N, Kulalert W, Walsh RM, Khalil A, Rheinwald JG, Hochedlinger K. 2009. Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature* **460**: 1145-1148.
- van Attikum H, Gasser SM. 2009. Crosstalk between histone modifications during the DNA damage response. *Trends Cell Biol* **19**: 207-217.
- Voo KS, Carlone DL, Jacobsen BM, Flodin A, Skalnik DG. 2000. Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol Cell Biol* **20**: 2108-2121.
- Waddington CH. 1942. The epigenotype. *Endeavour* **1**: 18-20.
- Wakayama T, Perry AC, Zuccotti M, Johnson KR, Yanagimachi R. 1998. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**: 369-374.
- Wang T, Chen K, Zeng X, Yang J, Wu Y, Shi X, Qin B, Zeng L, Esteban MA, Pan G et al. 2011. The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell Stem Cell* **9**: 575-587.
- Ware CB, Wang L, Mecham BH, Shen L, Nelson AM, Bar M, Lamba DA, Dauphin DS, Buckingham B, Askari B et al. 2009. Histone deacetylase inhibition elicits an evolutionarily conserved self-renewal program in embryonic stem cells. *Cell Stem Cell* **4**: 359-369.

- Wernig M. 2004. Functional integration of embryonic stem cell-derived neurons in vivo. *J Neurosci* **24**: 5258.
- Wernig M, Meissner A, Cassady JP, Jaenisch R. 2008. c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cell* **2**: 10-12.
- Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, Bernstein BE, Jaenisch R. 2007. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**: 318-324.
- Wiblin AE, Cui W, Clark AJ, Bickmore WA. 2005. Distinctive nuclear organisation of centromeres and regions involved in pluripotency in human embryonic stem cells. *J Cell Sci* **118**: 3861.
- Williams RRE, Azuara V, Perry P, Sauer S, Dvorkina M, Jorgensen H, Roix J, McQueen P, Misteli T, Merckenschlager M et al. 2006. Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. *J Cell Sci* **119**: 132-140.
- Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KH. 1997. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**: 810-813.
- Wilusz JE, Sunwoo H, Spector DL. 2009. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* **23**: 1494-1504.
- Woodcock CL. 2006. Chromatin architecture. *Curr Opin Struct Biol* **16**: 213-220.
- Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. 2011a. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev* **25**: 679-684.
- Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. 2011b. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**: 389-393.
- Wu JI, Lessard J, Crabtree GR. 2009. Understanding the words of chromatin regulation. *Cell* **136**: 200-206.
- Wu SC, Zhang Y. 2010. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* **11**: 607-620.
- Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P et al. 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**: 86-90.
- Xie H, Ye M, Feng R, Graf T. 2004. Stepwise reprogramming of B cells into macrophages. *Cell* **117**: 663-676.

- Yamanaka S. 2009. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460**: 49-52.
- Yamanaka S, Blau HM. 2010. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465**: 704-712.
- 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**: 1155-1165.
- Yang PK, Kuroda MI. 2007. Noncoding RNAs and intranuclear positioning in monoallelic gene expression. *Cell* **128**: 777-786.
- Young RA. 2011. Control of the embryonic stem cell state. *Cell* **144**: 940-954.
- Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R et al. 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**: 1917-1920.
- Yuan P, Han J, Guo G, Orlov YL, Huss M, Loh YH, Yaw LP, Robson P, Lim B, Ng HH. 2009. Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes Dev* **23**: 2507-2520.
- Zaratiegui M, Irvine DV, Martienssen RA. 2007. Noncoding RNAs and gene silencing. *Cell* **128**: 763-776.
- Zhao XY, Li W, Lv Z, Liu L, Tong M, Hai T, Hao J, Guo CL, Ma QW, Wang L et al. 2009. iPS cells produce viable mice through tetraploid complementation. *Nature* **461**: 86-90.